OM of: US-09-303-518D-131 to: SPTREMBL_19:* out_format : pfs

Date: Jun 30, 2002 8:05 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142714_4338/app_query.fasta_1.23501
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_1967 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-131
Query length: 1344
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 883.700000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_bacteria:Q44130	+	1583.00	2203.72	4.7e-115	449	Q44130 actinobacillus pleuropne
sp_bacteria:Q9X6S0	+	648.00	894.49	3.9e-42	451	Q9x6s0 porphyromonas gingivalis
sp_human:Q9UQ36	+	141.00	173.79	0.0194	1275	Q9uq36 homo sapiens (human). r
sp_human:O15038	+	141.00	170.31	0.0217	1783	O15038 homo sapiens (human). k
sp_human:O60382	+	141.00	170.27	0.0217	1791	O60382 homo sapiens (human). k
sp_human:Q9UQ35	+	141.00	165.81	0.0250	2752	Q9uq35 homo sapiens (human). r
sp_bacteriap:Q9WY86	+	135.00	176.18	0.0403	451	Q9wy86 thermotoga maritima. ele
sp_fungi:Q04051	-	128.00	152.00	0.2244	1802	Q04051 saccharomyces cerevisia
sp_human:Q9UHA8	+	127.00	148.08	0.2910	2296	Q9uha8 homo sapiens (human). s
sp_fungi:Q9C105	-	125.50	152.41	0.3103	1236	Q9c105 schizosaccharomyces pom
sp_rodent:Q99K60	-	122.00	154.49	0.4655	631	Q99k60 mus musculus (mouse). mu
sp_human:Q9UF83	-	121.50	154.67	0.4952	580	Q9uf83 homo sapiens (human). hy
sp_fungi:O94317	-	121.00	154.82	0.5271	534	O94317 schizosaccharomyces pomb
sp_human:Q9UQ39	+	118.50	145.28	1.00	956	Q9uq39 homo sapiens (human). rr
sp_human:Q9UQ40	+	118.50	142.40	1.10	1262	Q9uq40 homo sapiens (human). r
sp_plant:Q9SUV0	+	118.00	145.71	1.06	857	Q9suv0 arabidopsis thaliana (mo
sp_mammal:Q29071	-	116.00	147.94	1.29	528	Q29071 sus scrofa (pig). gastri
sp_plant:Q9LWY3	+	115.50	143.75	1.58	739	Q9lwy3 oryza sativa (rice). hyp
sp_invertebrate:Q23587	-	115.00	126.89	2.89	3507	Q23587 caenorhabditis elegan
sp_invertebrate:Q23804	+	114.50	142.21	1.89	749	Q23804 chironomus tentans (mi
sp_rvirus:Q98WW2	+	114.50	134.97	2.39	1504	Q98ww2 avian leukosis virus. g
sp_invertebrate:Q9BKV7	-	113.00	134.19	3.00	1325	Q9bkv7 leishmania major. ppg
sp_plant:O23054	-	112.50	145.87	2.21	402	O23054 arabidopsis thaliana (mo
sp_bacteria:O06470	+	112.00	145.09	2.42	405	O06470 burkholderia cepacia (ps
sp_virus:Q98201	+	112.00	141.38	2.73	579	Q98201 molluscum contagiosum vi
sp_bacteriap:Q9RYP2	+	111.50	145.16	2.58	376	Q9ryp2 deinococcus radiodurans
sp_invertebrate:Q9N5K0	-	111.50	133.09	3.80	1203	Q9n5k0 caenorhabditis elegan
sp_bacteria:P94909	+	110.50	140.22	3.46	529	P94909 microbacterium ammoniaph
sp_human:Q9H9M1	-	109.00	136.16	4.83	639	Q9h9m1 homo sapiens (human). nt
sp_bacteria:Q93LB1	+	108.50	143.34	4.11	299	Q93lb1 bifidobacterium animalis
sp_bacteria:Q9KXM5	+	108.50	133.10	5.69	802	Q9kxm5 streptomyces coelicolor
sp_human:Q9Y2N2	-	108.50	128.13	6.67	1294	Q9y2n2 homo sapiens (human). w
sp_invertebrate:P91365	-	108.50	122.47	7.99	2232	P91365 caenorhabditis elegan
sp_bacteria:Q9RFF2	+	108.00	135.96	5.56	569	Q9rff2 rhodobacter sphaeroides
sp_invertebrate:Q9U0V2	-	108.00	121.89	8.71	2207	Q9u0v2 leishmania major. pos
sp_invertebrate:Q9VXM5	-	108.00	121.00	8.96	2406	Q9vxm5 drosophila melanogast
sp_human:Q9H195	-	107.00	129.79	7.75	901	Q9h195 homo sapiens (human). mu
sp_rodent:Q9QY35	-	107.00	121.62	10.05	1979	Q9qy35 mus musculus (mouse). h
sp_human:Q14395	-	106.50	135.10	6.99	505	Q14395 homo sapiens (human). mu
sp_bacteriap:Q98K15	+	106.50	134.90	7.04	515	Q98k15 rhizobium loti (mesorhiz
sp_bacteria:Q46767	+	106.50	131.25	7.91	732	Q46767 escherichia coli. orf
sp_bacteriap:O51099	+	106.50	130.68	8.05	773	O51099 borrelia burgdorferi
sp_invertebrate:O61228	-	106.50	130.20	8.18	810	O61228 tenebrio molitor (y
sp_human:Q96HA1	-	106.50	128.18	8.72	984	Q96ha1 homo sapiens (human).
sp_invertebrate:Q95SX4	-	106.00	132.07	8.24	632	Q95sx4 drosophila melanoga
sp_invertebrate:Q960L2	-	106.00	129.55	8.93	806	Q960l2 drosophila melanoga
sp_invertebrate:P91495	+	106.00	120.94	11.76	1847	P91495 caenorhabditis ele
sp_invertebrate:Q9VXE6	-	106.00	119.67	12.25	2088	Q9vxe6 drosophila melanog
sp_fungi:Q96WV6	-	106.00	112.99	15.16	3971	Q96wv6 schizosaccharomyces
sp_plant:Q9SCB5	+	105.00	133.15	9.11	498	Q9scb5 lycopersicon esculent
sp_human:Q99552	-	105.00	112.70	17.51	3570	Q99552 homo sapiens (human)
sp_bacteria:Q9ZA63	+	104.50	138.80	8.14	270	Q9za63 neisseria meningitidi
sp_human:Q9UKW9	-	104.50	123.17	13.41	1217	Q9ukw9 homo sapiens (human)
sp_virus:Q83034	+	104.50	112.28	18.98	3473	Q83034 rice tungro spherica
sp_bacteria:Q93RW3	+	104.00	133.87	10.19	406	Q93rw3 streptomyces coelicol
sp_organelle:Q9MRR6	+	104.00	132.04	10.80	484	Q9mrr6 clethra arborea. atp
sp_invertebrate:O76894	-	104.00	118.44	16.68	1795	O76894 drosophila melanog
sp_bacteria:Q9EVN5	+	103.50	135.35	10.40	329	Q9evn5 pseudomonas stutzeri
sp_virus:Q91TW8	-	103.50	132.84	11.27	419	Q91tw8 maize rayado fino vir
sp_bacteriap:Q98J69	+	103.50	125.71	14.15	833	Q98j69 rhizobium loti (mesor
sp_virus:Q66537	-	103.50	124.82	14.55	907	Q66537 human herpesvirus 4 (
sp_invertebrate:Q9N8G4	-	103.50	121.50	16.18	1249	Q9n8g4 trypanosoma brucei
sp_bacteria:Q48926	+	103.50	113.13	21.14	2796	Q48926 mycobacterium bovis.
sp_organelle:Q9MRR2	+	103.00	130.71	12.90	481	Q9mrr2 carpenteria californi
sp_human:Q9BUS4	+	103.00	130.35	13.05	498	Q9bus4 homo sapiens (human).
sp_archeap:Q9HRT5	+	103.00	129.15	13.56	559	Q9hrt5 halobacterium sp. (st
sp_archeap:Q97A73	+	103.00	128.55	13.82	592	Q97a73 thermoplasma volcaniu
sp_human:Q14881	-	103.00	128.04	14.05	622	Q14881 homo sapiens (human).
sp_rodent:Q61191	-	103.00	115.68	20.85	2045	Q61191 mus musculus (mouse)
sp_rodent:Q9QWH2	-	103.00	115.68	20.85	2045	Q9qwh2 mus musculus (mouse)
sp_bacteria:Q93SF8	+	102.50	127.76	15.16	597	Q93sf8 eubacterium acidamino
sp_vertebrate:Q9IBC7	-	102.50	125.95	16.07	711	Q9ibc7 xenopus laevis (afric
sp_invertebrate:Q9NEB8	+	102.50	119.25	19.90	1356	Q9neb8 leishmania major.
sp_mammal:Q29070	-	102.00	133.63	13.45	317	Q29070 sus scrofa (pig). gas
sp_virus:Q9DWA5	+	102.00	125.67	17.34	683	Q9dwa5 rat cytomegalovirus (
sp_archeap:Q9YFA1	+	102.00	124.71	17.88	749	Q9yfa1 aeropyrum pernix. hyp
sp_invertebrate:O01576	-	102.00	123.96	18.32	805	O01576 caenorhabditis eleg
sp_human:Q14163	+	102.00	122.31	19.31	944	Q14163 homo sapiens (human).
sp_mammal:O18758	-	101.50	94.15	50.74	13288	O18758 sus scrofa domestic
sp_bacteriap:Q9A4U7	+	101.00	124.72	20.46	654	Q9a4u7 caulobacter crescentu
sp_bacteriap:Q98DA7	+	101.00	124.10	20.87	694	Q98da7 rhizobium loti (mesor
sp_bacteriap:Q97PQ2	+	101.00	121.24	22.86	914	Q97pq2 streptococcus pneumon
sp_bacteriap:P74275	+	101.00	120.41	23.47	990	P74275 synechocystis sp. (st
sp_human:Q9UHA8	-	101.00	111.68	31.02	2296	Q9uha8 homo sapiens (human)
sp_bacteriap:Q97G36	+	100.50	130.45	18.23	352	Q97g36 clostridium acetobuty
sp_bacteriap:Q9K7W6	+	100.50	130.10	18.43	364	Q9k7w6 bacillus halodurans.
sp_archeap:Q96ZX6	+	100.50	122.73	23.32	740	Q96zx6 sulfolobus tokodaii.
sp_bacteriap:Q9HXA4	+	100.50	122.20	23.72	779	Q9hxa4 pseudomonas aeruginos
sp_bacteriap:Q9AA67	+	100.50	117.33	27.71	1245	Q9aa67 caulobacter crescent
sp_invertebrate:O18182	+	100.50	116.47	28.48	1353	O18182 caenorhabditis ele
sp_fungi:O13308	+	100.50	114.89	29.96	1576	O13308 candida albicans (ye
sp_fungi:O74209	+	100.50	114.89	29.96	1576	O74209 candida albicans (ye
sp_bacteria:O67995	-	100.00	126.72	21.96	471	O67995 pseudomonas fluoresce
sp_bacteria:Q936X7	+	100.00	125.30	22.98	540	Q936x7 pseudomonas sp. (stra
sp_bacteria:O86842	+	100.00	124.91	23.27	561	O86842 streptomyces coelicol
sp_bacteriap:P73758	+	100.00	124.67	23.45	574	P73758 synechocystis sp. (st
sp_mammal:O19115	-	100.00	124.18	23.82	602	O19115 oryctolagus cuniculus
sp_human:Q9H088	+	100.00	119.56	27.60	939	Q9h088 homo sapiens (human).
sp_invertebrate:O17296	+	100.00	118.71	28.36	1019	O17296 caenorhabditis ele
sp_bacteria:O69806	+	99.50	135.39	17.81	191	O69806 streptomyces coelicol

seq_name: sp_bacteria:Q44130

seq_documentation_block:
ID Q44130 PRELIMINARY; PRT; 449 AA.
AC Q44130;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE 48 KDA OUTER MEMBRANE PROTEIN.
GN AOPA.
OS Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

```
OC   Actinobacillus.
OX   NCBI_TaxID=715;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ISU158;
RX   MEDLINE=96110919; PubMed=8557378;
RA   Cruz W.T., Nedialkov Y.A., Thacker B.J., Mulks M.H.;
RT   "Molecular characterization of a common 48-kilodalton outer membrane
RT   protein of Actinobacillus pleuropneumoniae.";
RL   Infect. Immun. 64:83-90(1996).
DR   EMBL; U24492; AAC43631.1; -.
SQ   SEQUENCE   449 AA;  48604 MW;  62CBDD17A8435418 CRC64;


alignment_scores:
            Quality: 1583.00              Length:    450
              Ratio:  4.059                Gaps:      3
 Percent Similarity:  86.667   Percent Identity: 68.889

alignment_block:
 US-09-303-518D-131 x Q44130   ..

 Align seg 1/1  to: Q44130  from: 1  to: 449

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         ||||||  ||||||||||||||||:::||||||||||||||||   |||
       1 MetIleThrIleLysLysGlyLeuAspLeuProIleAlaGlyThrProAl 17

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCG 100
         |||||||||:::::|||   :::::::::|||||||||:::|||||||
      17 aGlnValIleHisAsnGlyAsnThrValAsnGluValAlaMetLeuGlyG 34

     101 AAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCC 150
         ||||||||||||||||||||||||||||||:::::|||||:::::
      34 luGluTyrValGlyMetArgProSerMetLysValArgGluGlyAspVal 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGT 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 ATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAA 250
         ||||||||||||||||||||   :::::::::|||:::||||||||||
      67 lPheThrAlaProAlaSerGlyThrValValThrIleAsnArgGlyGluL 84

     251 AGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||   |||||||||:::::::|||
      84 ysArgValLeuGlnSerValValIleLysValGluGlyAspGluGlnIle 100

     301 GAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
            |||   ||||||              ||||||:::||||||:::|||::
     101 ThrPheThrArgTyrGluAlaAlaGlnLeuAlaSerLeuSerAlaGluGl 117

     351 AGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCC 400
         :|||::::::|||||||||:::|||||||||||||||||||:::|||||||
     117 nValLysGlnAsnLeuIleGluSerGlyLeuTrpThrAlaPheArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||:::||||||:::||||||   |||   :::||||||
     134 rgProPheSerLysValProAlaLeuAspAlaIleProSerSerIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCAT 500
         |||||||||||||||||||||||||||||||||||||   |||:::::
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProGluValValLe 167

     501 CAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCC 550
         :||||||         |||||||||   ||||||   ||||||:::||||
     167 uLysGluTyrGluThrAspPheLysAspGlyLeuThrValLeuThrArgL 184

     551 TG......ACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGAC 594
         ||         ::::::   :::::::::||||||   |||   ::::::
     184 euPheAsnGlyGlnLysProValTyrLeuCysLysAspAlaAspSerAsn 200

     595 GTGCCG...TCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGG 641
         ::::|||   |||            :::|||         :::|||:::||
     201 IleProLeuSerProAlaIleGluGlyIleThrIleLysSerPheSerGl 217

     642 CCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAG 691
         |   ||||||||||||||   ||||||||||||||||||:::::|||||
     217 yValHisProAlaGlyLeuValGlyThrHisIleHisPheValAspProV 234

     692 TCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCT 741
         ||||||||:::|||   ||||||   :::|||||||||||||||||||||
     234 alGlyAlaThrLysGlnValTrpHisLeuAsnTyrGlnAspValIleAla 250

     742 ATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGC 791
         ||||||:::||||||:::||||||:::|||   |||:::|||:::::::::
     251 IleGlyLysLeuPheThrThrGlyGluLeuPheThrAspArgIleIleSe 267

     792 CTTGGGCGGCCTGCAAGTCAACAAACCGCGCGCCTCTTGCGTACCGTTTTGG 841
         :|||:::|||   ||||||:::::|||||||||:::||||||   ||||
     267 rLeuAlaGlyProGlnValLysAsnProArgLeuValArgThrArgLeuG 284

     842 GTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAAC 891
         |||||:::::::|||||||||||||||||:::||||||      ::::::|||
     284 lyAlaAsnLeuSerGlnLeuThrAlaAsnGluLeuAsnAlaGlyGluAsn 300

     892 CGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCA 941
         ||||||||||||||||||||||||||:::||||||   |||   |||
     301 ArgValIleSerGlySerValLeuSerGlyAlaThrAlaAlaGlyProVa 317

     942 TGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCC 991
         ||||||||||||||||||      |||:::||||||:::   |||||||
     317 lAspTyrLeuGlyArgTyrAlaLeuGlnValSerValLeuAlaGluGlyA 334

     992 GCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCC 1041
         ||:::|||||||||||||||||||:::   |||      ||||||:::|||
     334 rgGluLysGluLeuPheGlyTrpIleMetProGlySerAspLysPheSer 350

    1042 ATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTT 1091
         ||||||||||||:::||||||||||||   |||   |||||||||:::||
     351 IleThrArgThrValLeuGlyHisPheGlyLys...LysLeuPheAsnPh 366

    1092 CACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTT 1141
         ||||||||||||:::||||||:::||||||||||||||||||||||:::|
     366 eThrThrAlaValHisGlyGlyGluArgAlaMetValProIleGlyAlaT 383

    1142 ATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGAT 1191
         ||||||||||||||||||||||||||:::||||||||||||||||||||
     383 yrGluArgValMetProLeuAspIleIleProThrLeuLeuLeuArgAsp 399

    1192 TTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATT 1241
         |||   :::||||||||||||||||||||   ||||||||||||||||
     400 LeuAlaAlaGlyAspThrAspSerAlaGlnAsnLeuGlyCysLeuGluLe 416

    1242 GGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACG 1291
         ||||||||||||||||||||||||||||:::::||||||||||||||   :
     416 uAspGluGluAspLeuAlaLeuCysThrTyrValCysProGlyLysAsnA 433

    1292 AATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAGAAGGAAGGC 1341
         ::||||||||:::||||||   :::||||||   ||||||||||||||||
     433 snTyrGlyProMetLeuArgAlaAlaLeuGluLysIleGluLysGluGly 449

seq_name: sp_bacteria:Q9X6S0

seq_documentation_block:
ID   Q9X6S0      PRELIMINARY;      PRT;   451 AA.
AC   Q9X6S0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   50 KDA ANTIGEN PG1.
OS   Porphyromonas gingivalis (Bacteroides gingivalis).
OC   Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OC   Porphyromonas.
OX   NCBI_TaxID=837;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=W50;
RA   Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA   Hocking D., Webb E.;
RT   "Porphyromonas gingivalis polypeptides and nucleic acids.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF144076; AAD33930.1; -.
SQ   SEQUENCE   451 AA;  49825 MW;  3E4CC9E66FE5AD74 CRC64;


alignment_scores:
            Quality:  648.00              Length:    452
              Ratio:  2.189                Gaps:      7
 Percent Similarity:  65.487   Percent Identity: 34.071

alignment_block:
 US-09-303-518D-131 x Q9X6S0   ..

 Align seg 1/1  to: Q9X6S0  from: 1  to: 451

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
         :::||||||   ||||||||||||   |||   :::   |||:::|||
       4 ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 20

      51 GCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTC...GCGTTGCTTG 97
             :::          ||||||   :::   :::   |||::::::
      20 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValP 37

      98 GCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAA 147
           :::::::::   |||:::   |||::::::      :::   |||:::
      37 roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyAsp 53

     148 GCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGT 197
           |||:::   |||:::::|||:::::   |||      |||   ::
      54 LysValArgAlaGlySerAlaLeuMetHisHisLysAlaTyrProGluMe 70

     198 AGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCG 247
         :   ||||||:::|||:::||||||||:::::   |||:::::||||||
      70 tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 87

     248 AAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAA 297
           ||||||   :::   |||:::   :::      :::|||   :::|||
      87 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 103

     298 ATCGAG...TTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAG 344
           |||   |||            |||:::|||         ||||||::
     104 TyrGluSerProProValGlyAspProSerAla.....LeuSerAl 117

     345 CGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTC 394
         :|||:::::::::::::   |||::::::||||||:::|||     ::::
     117 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 134

     395 GCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCC 444
         ::   ||||||::::::   :::   :::   |||      |||
     134 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 150

     445 ATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGT 494
         |||:::::::::|||      |||   |||||||||   |||
     151 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 167

     495 CATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGA 544
          |||::::::         :::   ::::::   ::::::   :::|||:
     167 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 184

     545 GCCGCCTGACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGAC 594
         :::::||||||      |||:::::::|||           |||:::?::
     184 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 200

     595 GTGCCGTCTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCC 644
         :::      :::||||||      :::|||:::|||         |||||
     201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 214

     645 GCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCG 694
           ||||||||||||      |||:::   |||:::      :::|||::::
     214 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 231

     695 GCGCGAATAAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATC 744
         ::   :::::|||||||||||||:::::      |||:::|||:::|||
     231 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 247

     745 GGACGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTT 794
         ||||||:::::::::|||||||:::   :::      |||:::::|||::
     248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 264

     795 GGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTG 844
         :   |||   ::::::                  :::|||   :::   |||:
     264 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 281

     845 CGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTT......GACGCGGAC 888
         :::::|||               |||:::|||:::      ::::::
     281 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 297

     889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGC 938
         :::|||||||||:::|||:::||||||:::|||
     298 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 314

     939 GCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
          :::   :::|||:::         :::||||||:::||||||   ||||
     314 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 331

     989 GCCGCAGC...AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
         ||   :::      ||||||||||||||:::||||||:::   |||:::
     331 lyAspAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 347

    1036 TACTCCATCACGCGCACCACTCTCGGCCATTTCCTA...AAAAACAAACT 1082
         ||||||::::::|||:::   ::::::   :::      |||||||||
     348 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysGl 364

    1083 CTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGA 1132
          :::   :::   :::   ::::::||||||:::||||||||||:::
     364 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 381

    1133 TCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTT 1182
           :::   |||:::||||||:::|||:::||||||   |||      |||
     381 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397

    1183 TTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTG 1232
         |||:::   :::|||:::   |||   |||:::   :::   ||||||
     398 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyIl 414

    1233 CTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGG 1282
           |||:::      ||||||:::|||   |||:::||||||      :
     414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 431

    1283 GCAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCTGGAAACCATTGAG 1332
         ::|||   |||         ::::::|||:::   |||:::   :::
     431 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 447

    1333 AAGGAA 1338
         ||||||
     448 LysGlu 449

seq_name: sp_human:Q9UQ36

seq_documentation_block:
ID   Q9UQ36      PRELIMINARY;      PRT;  1275 AA.
AC   Q9UQ36;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA BINDING PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016091; BAA83717.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1275 AA;  136869 MW;  45C2B2F85E98A6F6 CRC64;


alignment_scores:
            Quality:  141.00              Length:    446
              Ratio:  0.701                Gaps:     17
 Percent Similarity:  45.067   Percent Identity: 25.336

alignment_block:
 US-09-303-518D-131 x Q9UQ36   ..

 Align seg 1/1  to: Q9UQ36  from: 1  to: 1275

      32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
         |||   |||      ||||||   |||      :::   :::|||
     244 ProProArgHisArgArgSerProSerValSerSerProGluProAlaGl 260

      82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAA 131
          |||||||||         ::::::   ||||||   :::|||||||   |
     260 uLysSerArgSerSerArgArgArgArgSerAlaSerSerProArgThrL 277

     132 AATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACA 181
         ||:::                                          |||
     277 ysThr.........................................Thr 279

     182 AAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCC 231
         :::|||    ||||||
     280 SerArg.ArgGlyArg............................... 284

     232 GCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
           ||||||      |||   |||   ::::::::::|||:::   |||
     285 .....SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 297

     282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGG 331
         |||:::||||||:::::::::||||||   |||
     298 SerArgSerArgArgGluLysThrArgThrThrArgArg.......... 310

     332 CAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTA 381
             ::::::|||:::::::::   |||      |||   |||
     311 .......ArgAspArgSerGlySerSerGlnSerThrSerArgArg.... 323

     382 TGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGC 431
                                 :::|||:::          |||
     324 ..........................ArgGlnArgSerArgSerArgSer 331

     432 CGAGCCGTTCGCCATCTTCGTCAATGCGATGGA...CACCAATCCGCTGG 478
         |||::::::|||:::   |||      :::|||   |||:::   :::
     332 ArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerArgSerPr 348

     479 CTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGA.........AGAC 519
          :::|||      :::   :::      |||||||||         |||
     348 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgS 365

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
         :::|||   |||::::::   :::                   :::|||
     365 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 381

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATA 619
         :::      :::      |||   ||||||
     382 AlaProTrpLysArgSerArgSerArgAla................... 391

     620 TCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACG 669
                             ||||||:::               |||
     392 ..................SerProAlaThr..............HisA 397

     670 CACATTCATTTCATCGA............GCCAGTCGGCGCGAATAAAAC 707
           :::|||      |||            :::   ||||||:::   :::

     397 rgArgSerArgSerArgThrProLeuIleSerArgArgArgSerArgSer 413

     708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCG 757
         |||::::::           |||||||:::      ||||||   |||
     414 ArgThrSerProValSerArgArgArgSerArgSerArgThrSerValTh 430

     758 TAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAA 807
          :::|||   |||:::   ||||||               |||   :::
     430 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSerA 447

     808 GTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCA 857
           ::::::|||   |||::::::   |||   :::   :::   :::
     447 rgSerArgThrProProValThrArgArgArgSerArgSerArgThrPro 463

     858 ACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTT 907
         |||   |||||||||       |||      |||
     464 ThrThrArgArgArgSerArgSerArgThrProPro.............. 475

     908 CGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGC 957
                                 |||||||||                |||
     476 ..................ValThrArgArgArgSerArgSerArgThrP 486

     958 TACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
           |||   :::      |||   |||   |||   :::      :::
     486 roProValThrArgArgArgSerArgSerArgThrSerProIleThrArg 502

    1008 CGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTC 1057
         |||   :::   :::   ::::::   :::   ::::::      :::||
     503 ArgArgSerArgSerArgThrSerProValThrArgArgArgSerArgSe 519

    1058 TCGG......CCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCC 1101
         ||||      |||      ::::::::::::   :::      :::
     519 rArgThrSerProValThrArgArgArgSerArgSerArgThrSerProV 536

    1102 GTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGT 1151
              |||||||||   |||      |||         :::      |||
     536 alThrArgArgArgSerArgSerArgThrProProAlaIleArgArgArg 552

    1152 AATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCG 1201
         :::            |||                          :::   |||::
     553 SerArgSerArgThrProLeu..................LeuProArgLy 563

    1202 GCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGA...A 1248
         :|||   |||:::|||:::   :::::::::         :::|||   |
     563 sArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 580

    1249 GAAGACCTCGCTTTGTGCAG..............CTTCGTCTGCCCGGG 1283
         ||   ||||||   ::::::                    |||   |||
     580 rgThrProArgThrAlaArgGlyLysArgSerLeuThrArgSerProPro 596

    1284 CAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCT 1319
            ||||||   |||   |||   ::::::|||:::
     597 AlaIleArgArgArgSerAlaSerGlySerSerSer 608

seq_name: sp_human:O15038

seq_documentation_block:
ID   O15038      PRELIMINARY;      PRT;  1783 AA.
AC   O15038;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0324 PROTEIN (FRAGMENT).
GN   KIAA0324.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
```

```
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
DR   EMBL; AB002322; BAA20782.2; -.
FT   NON_TER      1      1
SQ   SEQUENCE   1783 AA;  190940 MW;  660302F6FD4179AB CRC64;


alignment_scores:
            Quality:  141.00              Length:    446
              Ratio:  0.701                Gaps:     17
 Percent Similarity:  45.067   Percent Identity: 25.336

alignment_block:
 US-09-303-518D-131 x O15038   ..

 Align seg 1/1  to: O15038  from: 1  to: 1783

      32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
         |||   |||      ||||||  |||      :::   :::|||
     752 ProProArgHisArgArgSerProSerValSerSerProGluProAlaGl 768

      82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAA 131
         |||||||||          ::::::  ||||||  :::|||||| |
     768 uLysSerArgSerSerArgArgArgArgSerAlaSerSerProArgThrL 785

     132 AATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACA 181
         ||:::                                        |||
     785 ysThr.....................................Thr 787

     182 AAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCC 231
         :::|||    ||||||
     788 SerArg.ArgGlyArg.............................. 792

     232 GCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
             ||||||      |||   |||   ::::::::::|||:::   |||
     793 .....SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 805

     282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGG 331
         |||:::||||||:::::::::||||||   |||
     806 SerArgSerArgArgGluLysThrArgThrThrArgArg..... 818

     332 CAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTA 381
               ::::::|||::::::::::   |||      |||   |||
     819 .......ArgAspArgSerGlySerSerGlnSerThrSerArgArg.... 831

     382 TGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGC 431
                                 :::|||:::          |||
     832 .........................ArgGlnArgSerArgSerArgSer 839

     432 CGAGCCGTTCGCCATCTTCGTCAATGCGATGGA...CACCAATCCGCTGG 478
         |||:::::::|||:::   |||      :::|||   |||:::   :::
     840 ArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerArgSerPr 856

     479 CTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGA.........AGAC 519
          :::|||      :::   :::      |||||||||          |||
     856 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgS 873

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
         :::|||   |||:::::   :::                   :::|||
     873 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 889

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATA 619
         :::      :::      |||   ||||||
     890 AlaProTrpLysArgSerArgSerArgAla..................... 899

     620 TCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACG 669
                              ||||||:::                |||
     900 ..................SerProAlaThr..............HisA 905

     670 CACATTCATTTCATCGA............GCCAGTCGGCGCGAATAAAAC 707

         :::|||      |||           :::   ||||||:::   :::
     905 rgArgSerArgSerArgThrProLeuIleSerArgArgArgSerArgSer 921

     708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCG 757
         |||::::::           ||||||:::      ||||||   |||
     922 ArgThrSerProValSerArgArgArgSerArgSerArgThrSerValTh 938

     758 TAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAA 807
          :::|||   |||:::   ||||||                  |||   :::
     938 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSerA 955

     808 GTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCA 857
         :::::|||   |||::::::   |||   :::   :::   :::
     955 rgSerArgThrProProValThrArgArgArgSerArgSerArgThrPro 971

     858 ACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTT 907
         |||   |||||||||          |||      |||
     972 ThrThrArgArgArgSerArgSerArgThrProPro.............. 983

     908 CGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGC 957
                                 |||||||||                |||
     984 ..................ValThrArgArgArgSerArgSerArgThrP 994

     958 TACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
         |||   :::      |||  |||   |||   :::      :::
     994 roProValThrArgArgArgSerArgSerArgThrSerProIleThrArg 1010

    1008 CGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTC 1057
         |||   :::   :::   ::::::   :::   ::::::      :::||
    1011 ArgArgSerArgSerArgThrSerProValThrArgArgArgSerArgSe 1027

    1058 TCGG......CCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCC 1101
         ||||      |||      ::::::::::::   :::      :::
    1027 rArgThrSerProValThrArgArgArgSerArgSerArgThrSerProV 1044

    1102 GTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGT 1151
             |||||||||   |||      |||          :::      |||
    1044 alThrArgArgArgSerArgSerArgThrProProAlaIleArgArgArg 1060

    1152 AATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCG 1201
         :::             |||                          :::   |||::
    1061 SerArgSerArgThrProLeu.................LeuProArgLy 1071

    1202 GCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGA...A 1248
         :|||   |||:::|||:::   :::::::::          :::|||   |
    1071 sArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 1088

    1249 GAAGACCTCGCTTTGTGCAG..............CTTCGTCTGCCCGGG 1283
         ||   ||||||   ::::::                  |||   |||
    1088 rgThrProArgThrAlaArgGlyLysArgSerLeuThrArgSerProPro 1104

    1284 CAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCT 1319
             ||||||   |||   |||   :::::::|||:::
    1105 AlaIleArgArgArgSerAlaSerGlySerSerSer 1116

seq_name: sp_human:O60382

seq_documentation_block:
ID   O60382      PRELIMINARY;      PRT;  1791 AA.
AC   O60382;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0324 (FRAGMENT).
GN   KIAA0324.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA   Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
```

```
RA   Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA   Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT   "Sequencing of Human Chromosome 16p13.3.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ricke D.O.;
RT   "Large Scale Sequence Analysis and Annotation with the Sequence
RT   Comparison Analysis (SCAN) System.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004493; AAC08453.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   1791 AA;  191306 MW;  3A7B5530AEE95F3E CRC64;


alignment_scores:
            Quality:  141.00              Length:    446
              Ratio:  0.701                Gaps:     17
 Percent Similarity:  45.067   Percent Identity: 25.336

alignment_block:
 US-09-303-518D-131 x O60382   ..

 Align seg 1/1  to: O60382  from: 1  to: 1791

      32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
         |||   |||      ||||||   |||      :::   :::|||
     753 ProProArgHisArgArgSerProSerValSerSerProGluProAlaGl 769

      82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCCTCGATGAA 131
         |||||||||          ::::::   ||||||   :::||||||   |
     769 uLysSerArgSerSerArgArgArgArgSerAlaSerSerProArgThrL 786

     132 AATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACA 181
         ||:::                                          |||
     786 ysThr.......................................Thr 788

     182 AAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCC 231
         :::|||    ||||||
     789 SerArg.ArgGlyArg................................ 793

     232 GCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
           ||||||      |||   |||   ::::::::::|||:::   |||
     794 .....SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 806

     282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGG 331
         |||:::||||||:::::::::||||||   |||
     807 SerArgSerArgArgGluLysThrArgThrThrArgArg.......... 819

     332 CAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTA 381
               ::::::|||:::::::::   |||      |||   |||
     820 .......ArgAspArgSerGlySerSerGlnSerThrSerArgArg.... 832

     382 TGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGC 431
                                  :::|||:::         |||
     833 ........................ArgGlnArgSerArgSerArgSer 840

     432 CGAGCCGTTCGCCATCTTCGTCAATGCGATGGA...CACCAATCCGCTGG 478
         |||::::::|||:::   |||      :::|||   |||:::   :::
     841 ArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerArgSerPr 857

     479 CTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGA.........AGAC 519
          :::|||      :::   :::      |||||||||          |||
     857 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgS 874

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
         :::|||   |||::::::   :::                    :::|||
     874 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 890

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATA 619
         :::      :::      |||   ||||||
     891 AlaProTrpLysArgSerArgSerArgAla.................... 900

     620 TCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACG 669
                                      |||||||:::              |||
     901 ..................SerProAlaThr.............HisA 906

     670 CACATTCATTTCATCGA............GCCAGTCGGCGCGAATAAAAC 707
           :::|||      |||           :::   |||||||:::   :::
     906 rgArgSerArgSerArgThrProLeuIleSerArgArgArgSerArgSer 922

     708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCG 757
         |||::::::           |||||||:::      ||||||   |||
     923 ArgThrSerProValSerArgArgArgSerArgSerArgThrSerValTh 939

     758 TAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAA 807
          :::|||   |||:::   ||||||                |||   :::
     939 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSerA 956

     808 GTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCA 857
         :::::::|||   |||::::::   |||   :::   :::   :::
     956 rgSerArgThrProProValThrArgArgArgSerArgSerArgThrPro 972

     858 ACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTT 907
         |||   |||||||||          |||      |||
     973 ThrThrArgArgArgSerArgSerArgThrProPro.............. 984

     908 CGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGC 957
                                      |||||||||              |||
     985 ..................ValThrArgArgArgSerArgSerArgThrP 995

     958 TACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
           |||   :::      |||   |||   |||   :::      :::
     995 roProValThrArgArgArgSerArgSerArgThrSerProIleThrArg 1011

    1008 CGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTC 1057
         |||   :::   :::   ::::::   :::   ::::::      ::::|
    1012 ArgArgSerArgSerArgThrSerProValThrArgArgArgSerArgSe 1028

    1058 TCGG......CCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCC 1101
         ||||      |||      ::::::::::::   :::      :::
    1028 rArgThrSerProValThrArgArgArgSerArgSerArgThrSerProV 1045

    1102 GTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGT 1151
               |||||||||   |||      |||           :::      |||
    1045 alThrArgArgArgSerArgSerArgThrProProAlaIleArgArgArg 1061

    1152 AATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCG 1201
         :::              |||                        :::   ||||::
    1062 SerArgSerArgThrProLeu.................LeuProArgLy 1072

    1202 GCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGA...A 1248
         :|||   |||:::|||:::   :::::::::          :::|||   |
    1072 sArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 1089

    1249 GAAGACCTCGCTTTGTGCAG..............CTTCGTCTGCCCGGG 1283
         ||   ||||||   ::::::                |||   |||
    1089 rgThrProArgThrAlaArgGlyLysArgSerLeuThrArgSerProPro 1105

    1284 CAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCT 1319
              ||||||   |||   |||   :::::::|||:::
    1106 AlaIleArgArgArgSerAlaSerGlySerSerSer 1117

seq_name: sp_human:Q9UQ35

seq_documentation_block:
ID   Q9UQ35      PRELIMINARY;      PRT;  2752 AA.
AC   Q9UQ35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA BINDING PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016092; BAA83718.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   2752 AA;  299672 MW;  109C64F181097123 CRC64;


alignment_scores:
           Quality:  141.00            Length:    446
             Ratio:  0.701              Gaps:     17
 Percent Similarity:  45.067   Percent Identity: 25.336

alignment_block:
 US-09-303-518D-131 x Q9UQ35   ..

 Align seg 1/1  to: Q9UQ35  from: 1  to: 2752

      32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
         |||   |||      ||||||   |||      :::   :::|||
    1721 ProProArgHisArgArgSerProSerValSerSerProGluProAlaGl 1737

      82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAA 131
          |||||||||         ::::::   ||||||   :::||||||   |
    1737 uLysSerArgSerSerArgArgArgArgSerAlaSerSerProArgThrL 1754

     132 AATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACA 181
         ||:::                                          |||
    1754 ysThr.....................................Thr 1756

     182 AAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCC 231
         :::|||    ||||||
    1757 SerArg.ArgGlyArg.............................. 1761

     232 GCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
              ||||||      |||   |||   ::::::::::|||:::   |||
    1762 .....SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 1774

     282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGG 331
            |||:::||||||:::::::::||||||   |||
    1775 SerArgSerArgArgGluLysThrArgThrThrArgArg........... 1787

     332 CAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTA 381
              ::::::|||:::::::::   |||      |||   |||
    1788 .......ArgAspArgSerGlySerSerGlnSerThrSerArgArg.... 1800

     382 TGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGC 431
                                  :::|||:::         |||
    1801 ........................ArgGlnArgSerArgSerArgSer 1808

     432 CGAGCCGTTCGCCATCTTCGTCAATGCGATGGA...CACCAATCCGCTGG 478
         |||:::::|||:::   |||      :::|||   |||:::   :::
    1809 ArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerArgSerPr 1825

     479 CTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGA.........AGAC 519
          :::|||      :::   :::      |||||||||         |||
    1825 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgS 1842

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
           :::|||   |||:::::   :::                   :::|||
    1842 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 1858

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATA 619
         :::      :::      |||   ||||||
    1859 AlaProTrpLysArgSerArgSerArgAla..................... 1868

     620 TCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACG 669
                              ||||||:::         |||
    1869 ..................SerProAlaThr.............HisA 1874

     670 CACATTCATTTCATCGA............GCCAGTCGGCGCGAATAAAAC 707
           :::|||      |||           :::   ||||||:::   :::
    1874 rgArgSerArgSerArgThrProLeuIleSerArgArgArgSerArgSer 1890

     708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCG 757
         |||::::::         ||||||:::     ||||||   |||
    1891 ArgThrSerProValSerArgArgArgSerArgSerArgThrSerValTh 1907

     758 TAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAA 807
          :::|||   |||:::   ||||||                |||   :::
    1907 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSerA 1924

     808 GTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCA 857
           :::::::|||   |||::::::   |||   :::   :::   :::
    1924 rgSerArgThrProProValThrArgArgArgSerArgSerArgThrPro 1940

     858 ACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTT 907
         |||   |||||||||         |||     |||
    1941 ThrThrArgArgArgSerArgSerArgThrProPro.............. 1952

     908 CGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGC 957
                                 |||||||||                |||
    1953 ..................ValThrArgArgArgSerArgSerArgThrP 1963

     958 TACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
           |||   :::      |||   |||   |||   :::      :::
    1963 roProValThrArgArgArgSerArgSerArgThrSerProIleThrArg 1979

    1008 CGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTC 1057
         |||   :::   :::   ::::::   :::   ::::::      :::||
    1980 ArgArgSerArgSerArgThrSerProValThrArgArgArgSerArgSe 1996

    1058 TCGG......CCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCC 1101
         ||||      |||      ::::::::::::   :::      :::
    1996 rArgThrSerProValThrArgArgArgSerArgSerArgThrSerProV 2013

    1102 GTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGT 1151
              |||||||||   |||      |||          :::      |||
    2013 alThrArgArgArgSerArgSerArgThrProProAlaIleArgArgArg 2029

    1152 AATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCG 1201
         :::            |||                        :::   |||::
    2030 SerArgSerArgThrProLeu.................LeuProArgLy 2040

    1202 GCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGA...A 1248
         :|||   |||:::|||:::   :::::::::         :::|||   |
    2040 sArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA 2057

    1249 GAAGACCTCGCTTTGTGCAG..............CTTCGTCTGCCCGGG 1283
         ||   ||||||   :::::                |||   |||
    2057 rgThrProArgThrAlaArgGlyLysArgSerLeuThrArgSerProPro 2073

    1284 CAAATACGAATACGGCCCGCTGTTGCGCAAAGTGCT 1319
            ||||||   |||   |||   ::::::|||:::
    2074 AlaIleArgArgArgSerAlaSerGlySerSerSer 2085

seq_name: sp_bacteriap:Q9WY86

seq_documentation_block:
ID   Q9WY86      PRELIMINARY;      PRT;   451 AA.
AC   Q9WY86;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ELECTRON TRANSPORT COMPLEX PROTEIN, PUTATIVE.
GN   TM0244.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
```

```
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001708; AAD35335.1; -.
DR   TIGR; TM0244; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001949; Complex1_51K.
DR   Pfam; PF01512; Complex1_51K; 1.
DR   Pfam; PF00037; fer4; 1.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Iron-sulfur; Complete proteome.
SQ   SEQUENCE   451 AA;  49515 MW;  7B2D7272CFC6E418 CRC64;


alignment_scores:
            Quality:  135.00              Length:    444
              Ratio:  0.603                Gaps:     17
 Percent Similarity:  50.450   Percent Identity: 20.270

alignment_block:
 US-09-303-518D-131 x Q9WY86   ..

 Align seg 1/1  to: Q9WY86  from: 1  to: 451

      49 GAGCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGG 98
         ::::::   |||      :::|||         :::|||   ::::::
      35 AspLysProIleGluArgAlaProLeuProGlnLysValPheValPheLe 51

      99 CGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAG 148
          :::::::::::|||         :::      ::::::   |||:::
      51 uSerAsnHisAlaGlyAsnProAlaLysProValValSerProGlyAspG 68

     149 CCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTA 198
           ||||||   ||||||||||:::   |||   :::        ::::::
      68 luValLysThrGlyGlnValIleGlyGluProGluGlyPheIleSerAla 84

     199 GTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGA 248
         :::   :::|||:::::|||:::::      |||
      85 TyrLeuHisSerProValThrGlyArgValLeuGluIle........... 97

     249 AAAGCGCGTACTTCAGTCAGTCGTG.................ATTGCCG 280
          |||::::::|||:::   ::::::                |||:::
      98 .LysLysIleLeuHisProIleLeuGlyLysProIleGluAlaIleVal I 114

     281 TTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTG 330
         ::|||   :::::::::   |||:::   :::             :::
     114 leGluArgThrSerAspAspGluTrpValHisIleGluThrGlyAspPhe 130

     331 GCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCA..... 375
         ::::::|||:::|||::::::   :::   :::   ::::::
     131 GluArgMetSerLysGluGluIleLeuGluIleIleLysLysAlaGlyIl 147

     376 ....GGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTG 421
             ||||||      |||:::               :::      |||
     147 eValGlyLeuGlyGlyAlaMetPheProThrHisValLysLeuSerProP 164

     422 CCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAAT 471
             :::   :::      :::::::::||||||:::   :::   :::
     164 roProGluLysLysValAspThrLeuIleValAsnGlyAlaGluCysGlu 180

     472 CCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTT 521
         |||:::   :::              ::::::   |||   |||||||||::
     181 ProValLeuThrIleAspHisArgLeuMetLeuGluArgAlaGluAspIl 197

     522 CAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATG 571
         :   :::|||:::|||::::::   ::::::            :::   |
     197 eLeuGlnGlyIleLeuIleMetMetLysVal...........LeuGlyV 210

     572 TGTGTAAAGCA...GCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAAT 618
         ||   ||||||   :::|||::::::      :::   :::|||   :::
     210 alGlnLysAlaValValGlyValGluSerAsnLysMetAspAlaTyrHis 226

     619 ATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCAC 668
                       |||   |||   :::|||:::   ::::::
     227 AsnLeuLysLysValPheLysGly...TyrProValAspValAlaLeuLe 242

     669 GCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCA 718
          :::      :::      |||   ||||||:::|||
     242 uArgThrLysTyr......ProGlnGlyAlaGluLys............. 252

     719 TCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGT 768
                 |||   :::   ||||||
     253 ........GlnLeuIleTyrAlaIle....................... 258

     769 CTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACC 818
               |||   |||:::|||      |||||||||   ::::::
     259 ......ThrGlyArgMetValProArgGlyGlyLeuProMetAspValGl 273

     819 GCGCCTCTTGCGT.........ACCGTTTTGGGTGCGAAGGTGTCTCAAC 859
             :::::::::          |||   :::::::::|||   :::   :
     273 yValValValGlnAsnValGlyThrCysValAlaValLysGluAlaValV 290

     860 TTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCG 909
         ::   :::      ||||||:::            |||   :::   :::
     290 alAspGlyLysProLeuValGluArgGlyMetThrValSerGlyAspAla 306

     910 GTATTGAAC.......................GGTGCGATTGCACAAGG 935
         |||   |||                     |||:::   ::::::
     307 ValLysAsnGlnLysAsnLeuIleValArgIleGlyThrProValLysAs 323

     936 CGCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAG 985
          :::   ||||||   |||                          |||:::|
     323 pValIleAspTyrCysGlyGly.....................IleAspG 333

     986 AAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
         ||:::   ::::::   ::::::|||      :::|||:::
     333 luAsnThrGluArgValIleLeuGly......GlyProMetMetGly... 346

    1036 TACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTT 1085
            |||||||||:::                 |||         :::::
     347 IleSerIleThrAsn.................LeuAspIleProValMe 357

    1086 CAAGTTCACGACAGCCGTCAACGGC.....................GGCG 1114
         :|||   |||:::::::::::::::                    :
     357 tLysGlyThrSerGlyIleThrAlaPheLeuProLysLysSerArgProG 374

    1115 ACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGAC 1164
         :::::      ::      ::         :::|||   |||::::::
     374 lnLysProCysIleArgCysSerGluCysValGlnValCysProMetAsn 390

    1165 ATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAG 1214
         :::   |||   ||||||                          |||::
     391 LeuGlnProTyrLeuLeuTyrLeuLeuSerThrLysArgLysTyrAspGl 407

    1215 CGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGT 1264
         :|||        |||   ::::::      |||      ::   |
     407 uAlaValGluAsnGlyLeuMetAspCysIleGlu......CysGlySerC 422

    1265 GCAGCTTCGTCTGCCCGGGCAAATACGAATAC 1296
         ||:::::::::|||||||:::|||   |||:::
     422 ysThrTyrThrCysProSerLysIleGluHis 432

seq_name: sp_fungi:Q04051

seq_documentation_block:
ID   Q04051      PRELIMINARY;      PRT;  1802 AA.
AC   Q04051;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
```

```
DE   HKR1P.
GN   HKR1 OR YDR420W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dietrich F.S.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
RA   Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA   Hunicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA   Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
RA   Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA   Winant A., Yelton M., Botstein D., Davis R.W.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Jia Y., Cherry J.M.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U33007; AAB64857.1; -.
DR   SGD; S0002828; HKR1.
SQ   SEQUENCE   1802 AA;  188945 MW;  0FF726990C7CDC51 CRC64;


alignment_scores:
            Quality:  128.00              Length:    467
              Ratio:  0.554                Gaps:     24
 Percent Similarity:  49.465   Percent Identity: 23.340

alignment_block:
 US-09-303-518D-131/rev x Q04051   ..

 Align seg 1/1  to: Q04051  from: 1  to: 1802

    1337 TCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTA 1288
          |||   ::::::||||||::::::      |||:::   |||||||||
     390 SerAspThrValValSerThrSerSerThrAsnThrValProTyrSerSe 406

    1287 TTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATT 1238
           :::      :::|||   :::|||   |||   ||||||||||:::
     406 rVal...HisSerThrPheValHis...AlaThrSerSerSerThrTyrI 421

    1237 CCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCG 1188
           ::::::            :::   |||:::|||   :::::::::|||
     421 leSerSerSerLeuTyrSerSerProSerLeuSerAlaSerValSerSer 437

    1187 CGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGT 1138
          :::   :::   :::         |||      |||:::   |||   ||
     438 HisPheGlyValAlaProPheProSerAlaTyrIleSerPheSerServa 454

    1137 GCCGATC.........GGTACCATGGCGCGGTCGCCGCCGTTGACGGCTG 1097
          |||:::         :::|||   :::   ||||||      :::::::|
     454 lProValAlaValSerSerThrTyrThrSerSerProSerAlaSerValV 471

    1096 TCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGC 1047
          |||||                        :::   |||   :::|||
     471 alValProSerAlaTyrAlaSerSerProSerValProValAlaValSer 487

    1046 GTGATGGAGTATTTGTCCGGCTGCGGCGCA..................... 1017
                |||   |||:::   :::|||
     488 ...SerThrTyrThrSerSerProSerAlaProAlaAlaIleSerSerTh 503

    1016 .......ACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGG 974
                 ::::::|||::::::   :::         |||:::   |||:
     503 rTyrThrSerSerProSerAlaProValAlaValSerSerThrTyrThrS 520

     973 AAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATC 924
          ::                                 ||||||:::||||||
     520 erSer.......................ProSerAlaProAlaAlaIle 528

     923 ...GCACCGTTCAATACCGAACCGGAA......ATCACGCGGTTGTCCGC 883
            :::   ::::::::::::|||:::      ::::::      |||::
     529 SerSerThrTyrThrSerSerProSerAlaProValAlaValSerSerTh 545

     882 GTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGG 833
          :   |||:::||||||:::                   ||||||   ::::
     545 rTyrThrSerSerProSer.................AlaProAlaAlaI 556

     832 TACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCG......CCCAAGGCA 789
          ::           :::      |||      |||      |||   |||
     556 le.........SerSerThrTyrThrSerSerProSerAlaProValAla 569

     788 ACCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGC 739
          ::::::   :::          |||        :::   |||:::||
     570 ValSerSerThrTyrThrSerSerPro.........SerAlaProValAl 583

     738 AATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCG.... 693
          ||||:::|||                  |||         :::|||
     583 aIleSerSer.................ThrTyrThrSerSerProSerV 594

     692 ........................ACTGGCTCGATGAAATGAATGTGC 669
                                  |||:::|||
     594 alProValAlaValSerSerThrTyrThrSerSer............... 605

     668 GTGCCACTCAAGCCGGCAGGA..............TGCGGGCCGCCAAA 634
            |||      ||||||:::                  :::   |||::
     606 ...ProSerAlaProAlaAlaIleSerSerThrTyrThrSerSerProSe 621

     633 TTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTG 584
          ::::   |||::::::   ::::::::::::::::::   ||||||||||
     621 rAlaProValAlaValSerSerThrTyrThrSerSerProSerAlaProA 638

     583 CTGCTTTACACACATGGATTTTACGTTCG.................GTC 552
          |||||:::   :::            |||                |||
     638 laAlaIleSerSerThrTyrThrSerSerProSerValProValAlaVal 654

     551 AGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTT 502
                   |||:::   |||        |||:::   |||:::::
     655 SerSerThrTyrThrSerSerPro.........SerAlaProAlaAlaIl 668

     501 GATGATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCATTGA 452
          :      |||      |||:::   |||      |||:::::::::   |
     668 eSerSerThrTyrThrSerSerProSerValProValAlaValSerSerT 685

     451 CGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGA 402
          ||      ::::::   ||||||   :::|||         :::::::::
     685 hrTyrThrSerSerProSerAlaProAlaAla.........IleSerSer 698

     401 CGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCAC 352
                |||:::         |||      :::   :::      ||
     699 ThrTyrThrSerSerProSerAlaPro......ValAlaValSerSerTh 713

     351 TTTTTCGCTGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACT 302
          |::::::             |||   |||:::            |||::::
     713 rTyrThrSerSerProSerAlaProAlaAla.........IleSerSerT 727

     301 CGATTTCGTCGTTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGC 252
          ::   :::|||   ||||||:::   ::::::::::::   :::
     727 hrTyrThrSerSerProSerAlaProValAlaValSerSerThrTyrThr 743

     251 TTTTCGCCACGGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAA 202
          ||||||          |||||||||            :::::::::   ::
     744 SerSerProSerAlaProAlaAlaIle.........SerSerThrTyrTh 757

     201 TACTACGCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGA 152
          :::::::|||:::       ::::::   :::|||||||:::         |
     757 rSerSerProSerAlaProValAlaValSerSerThrTyr.........T 771

     151 CGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCT 102
          ||:::|||||||||   :::   :::   :::          |||   |||
     771 hrSerSerProSerAlaLeuValValLeuSerSerThrSerThrSerSer 787
```

```
    101 ..................TCGCCAAGCAACGCGACTTCGGTAATGGCCGG 70
                           |||||||||:::   :::::::::   :::||
    788 ProTyrAspIleValTyrSerProSerThrPheAlaAlaIleSerSerGl 804

     69 G 69
        |
    804 y 804

seq_name: sp_human:Q9UHA8

seq_documentation_block:
ID   Q9UHA8      PRELIMINARY;      PRT;  2296 AA.
AC   Q9UHA8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SPLICING COACTIVATOR SUBUNIT SRM300.
GN   SRM300.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20132238; PubMed=10668804;
RA   Blencowe B.J., Bauren G., Eldridge A.G., Issner R., Nickerson J.A.,
RA   Rosonina E., Sharp P.A.;
RT   "The SRm160/300 splicing coactivator subunits.";
RL   RNA 6:111-120(2000).
DR   EMBL; AF201422; AAF21439.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   2296 AA;  251964 MW;  17C0BD4EA10A9CF9 CRC64;


alignment_scores:
            Quality:  127.00            Length:    418
              Ratio:  0.722               Gaps:     17
 Percent Similarity:  42.105   Percent Identity: 24.641

alignment_block:
 US-09-303-518D-131 x Q9UHA8   ..

 Align seg 1/1  to: Q9UHA8  from: 1  to: 2296

      32 CCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACC 81
         |||   |||      ||||||   |||      :::   :::|||
    1721 ProProArgHisArgArgSerProSerValSerSerProGluProAlaGl 1737

      82 GAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAA 131
          |||||||||         ::::::   ||||||   :::|||||| ·  |
    1737 uLysSerArgSerSerArgArgArgArgSerAlaSerSerProArgThrL 1754

     132 AATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACA 181
         ||:::                                        |||
    1754 ysThr.....................................Thr 1756

     182 AAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCC 231
         :::|||    ||||||
    1757 SerArg.ArgGlyArg.............................. 1761

     232 GCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGT 281
              ||||||      |||   |||   :::::::::|||:::   |||
    1762 .....SerProSerProLysProArg...GlyLeuGlnArgSer...Arg 1774

     282 TGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGG 331
             |||:::||||||::::::::||||||   |||
    1775 SerArgSerArgArgGluLysThrArgThrThrArgArg........... 1787

     332 CAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTA 381
                 ::::::|||:::::::::   |||      |||   |||
    1788 .......ArgAspArgSerGlySerSerGlnSerThrSerArgArg.... 1800

     382 TGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGC 431
                                        :::|||:::         |||
    1801 .........................ArgGlnArgSerArgSerArgSer 1808

     432 CGAGCCGTTCGCCATCTTCGTCAATGCGATGGA...CACCAATCCGCTGG 478
         |||:::::::|||:::   |||      :::|||   |||:::   :::
    1809 ArgValThrArgArgArgArgGlyGlySerGlyTyrHisSerArgSerPr 1825

     479 CTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGA.........AGAC 519
          :::|||      :::   :::     |||||||||         |||
    1825 oAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgS 1842

     520 TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCA 569
         :::|||   |||::::::   :::                    :::|||
    1842 erArgThrProProThrSerArgLysArgSerArgSerArgThrSerPro 1858

     570 TGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATA 619
         :::      :::      |||   ||||||
    1859 AlaProTrpLysArgSerArgSerArgAla..................... 1868

     620 TCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACG 669
                                    ||||||:::              |||
    1869 ..................SerProAlaThr.............HisA 1874

     670 CACATTCATTTCATCGA............GCCAGTCGGCGCGAATAAAAC 707
         :::|||      |||           :::   ||||||:::   :::
    1874 rgArgSerArgSerArgThrProLeuIleSerArgArgArgSerArgSer 1890

     708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCG 757
         |||::::::            |||||||::      ||||||   |||
    1891 ArgThrSerProValSerArgArgArgSerArgSerArgThrSerValTh 1907

     758 TAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAA 807
          :::|||   |||:::   ||||||                |||   :::
    1907 rArgArgArgSerArgSerArgAlaSerProValSerArgArgArgSerA 1924

     808 GTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCA 857
         :::::::|||   |||::::::   |||   :::   :::   :::
    1924 rgSerArgThrProProValThrArgArgArgSerArgSerArgThrPro 1940

     858 ACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTT 907
         |||   |||||||||         |||      |||
    1941 ThrThrArgArgArgSerArgSerArgThrProPro.............. 1952

     908 CGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGC 957
                                    |||||||||              |||
    1953 ..................ValThrArgArgArgSerArgSerArgThrP 1963

     958 TACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
           |||   :::            |||||||||   ::::::|||
    1963 roProValThr............ArgArgArgSerArgSerArg...... 1973

    1008 CGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTC 1057
                                                        ::
    1974 ..............................................Th 1974

    1058 TCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAAC 1107
         :   |||:::   ::::::::::::   :::      :::        |
    1974 rSerProIleThrArgArgArgSerArgSerArgThrSerProValThrA 1991

    1108 GGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCC 1157
         ||||||||   |||      |||
    1991 rgArgArgSerArgSerArgThr........................... 1998

    1158 GTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATA 1207
                                    :::      :::|||||||||
    1999 ........................SerProValThrArgArgArgSe 2006

    1208 CCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTC 1257
          |||:::|||::::::              :::   |||||||||   |
    2006 rArgSerArgThrSer............ProValThrArgArgArgSerA 2019

    1258 GC 1259
```

```
      ||
    2019 rg 2019

seq_name: sp_fungi:Q9C105

seq_documentation_block:
ID   Q9C105      PRELIMINARY;      PRT;  1236 AA.
AC   Q9C105;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE GLUCOAMYLASE I (ALPHA-1,4-GLUCAN GLUCOSIDASE), EXTRACELLULAR
DE   STARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STA1, CONTAINS
DE   CHITINASE FAMILY SIGNATURE.
GN   SPAPB1E7.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL590605; CAC36921.1; -.
DR   InterPro; IPR001917; Aminotransf_2.
DR   InterPro; IPR001579; Chitinase_2.
DR   Pfam; PF00192; chitinase_2; 2.
DR   PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
SQ   SEQUENCE   1236 AA;  123387 MW;  5A2D33A30B87CDD8 CRC64;


alignment_scores:
            Quality:  125.50            Length:    437
              Ratio:  0.570              Gaps:     17
 Percent Similarity:  50.343   Percent Identity: 24.714

alignment_block:
 US-09-303-518D-131/rev x Q9C105   ..

 Align seg 1/1  to: Q9C105  from: 1  to: 1236

    1343 CAGCCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTA 1294
         ||||||:::   |||:::::::||||||:::::          :::
     458 GlnProThrAlaSerValAlaSerSerSerValSerValProSerSerSe 474

    1293 TTCGTATTTGCCC...GGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTT 1247
          |||      |||   ::::::|||   :::           |||||||
     474 rSerValGlnProGlnSerSerThrProIle...........SerSerS 487

    1246 CGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACG 1197
         |||||:::::::::::::|||::::::   :::|||||||:::|||
     487 erSerSerAlaSerSerProGlnSer...ThrLeuSerThrSerSerGlu 502

    1196 ATTAAATCGCGCAAAAGCAAGGTA.........GGCAGGATGTCCAACGG 1156
         :::   |||:::   |||::::::         |||      :::
     503 ValValSerGluValSerSerThrLeuLeuSerGlySerSerAlaIlePr 519

    1155 CATTACGCGCTCATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGT 1106
           |||   |||   :::|||   :::::::::      |||   |||:
     519 oSerThrSerSerSerThrProSerSerSerIleIle...SerSerProM 535

    1105 TGACGGCTGTCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGA 1056
         ::|||:::|||
     535 etThrSerVal..................................... 538

    1055 GTGGTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAA 1006
                              ||||||:::   ::::::         ::
     539 ....................LeuSerSerSerSerSerIleProThrSe 548

    1005 CAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGC 956
         :||||||            ||||||||||||   |||

    548 rSerSerSerAspPheSerSerSerIleThrThrIle..... 560

    955 GTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAA 906
                 |||          :::      |||   :::   :::
    561 ........SerSerGlyIleSerSerSerSerIleProSerThrPheSer 574

    905 CCGGAAATCACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTG 856
                ::   ||||||:::||||||:::||||||:::   :::
    575 SerValSerSerIleLeuSerSerSerThrSerSerProSerSerThr.. 590

    855 AGACACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTT 806
                                                      :::
    591 ..............................................SerL 592

    805 GCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGTTACG 756
                   ::::::::::|||   |||:::|||           |||
    592 euSerIleSerSerSerSerThrSerSerThrPheSerSerAlaSerThr 608

    755 AACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACACGGT 706
        ::::::   |||   :::|||:::|||                   |||::
    609 SerSer...ProSerSerIleSerSerSerIleSerSerSerSerThrIl 624

    705 TTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCACTCAAGC 656
        :|||   :::||||||   |||   :::   :::   :::       :::
    624 eLeuSerSerProThrProSerThrSerSerLeuMetIleSerSerSerS 641

    655 CGGCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGCATTT 606
                   |||                 |||||||||::::::
    641 erIleIleSerGlySerSer...........SerIleLeuSerSerSer 653

    605 TCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTC 556
        :::                 ::::::|||   |||:::
    654 IleSerThrIleProIleSerSerSerLeuSerThrTyrSerSerSerVa 670

    555 GGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTT 506
         :::         :::|||            |||   ||||||::::::|
    670 lIleProSerSerSerThr...........LeuValSerSerSerSerS 683

    505 CTTTGATGATGACC.........GTAGGGTCGGCAGCCAGCGGATTGGTG 465
        |||||:::::::::         |||:::|||:::::::|||:::   :::
    683 erLeuIleValSerSerSerProValAlaSerSerSerSerSerProIle 699

    464 TCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGAT 415
              :::   :::   :::::::::   ||||||||   :::::::||
    700 ProSerSerSerSerLeuValSerThrTyrSerAlaSerLeuSerAsnIl 716

    414 TTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCA 365
        |      ::::::         :::|||:::   :::   :::   |||
    716 eThrHisSerSerLeuSerLeuThrAlaMetSerSerSerSerAlaIleP 733

    364 GGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGCTTCAGGTACG 315
                ::   |||   |||      ||||||:::|||:::|||
    733 roThrSerValAsnSerSerThrLeuIleThrAlaSerSerSerAsnThr 749

    314 TAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGGCAATCACGACTGA 265
              |||   ||||||:::|||      |||:::||||||::::::
    750 LeuLeuSer...SerIleThrSer......SerSerAlaIleValSer.. 762

    264 CTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCGATTTTGCCTGAAG 215
           ||||||      |||         |||::::::   ||||||:::|
    763 ....SerThrThrValSerAsn......IleSerSerAsnLeuProSerA 776

    214 CCGGCGCAGTAAATACTACG.........................CCC 192
        ||   |||   :::   :::
    776 laThrAlaSerSerGlnSerGlnLeuThrAsnSerSerThrLeuAlaThr 792

    191 GGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACC 142
        :::::::::||||||||:::|||:::      :::   |||:::|||
    793 SerLeuTyrLeuSerSerSerSerSerArgThrIleSerThrSerSerTh 809

    141 TTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCA 92
         :::            :::   :::   ||||||   ||||||   ::::
    809 rAsnGluTyrAsnThrSerPheHisAlaProThrValSerSerThrThrS 826
```

```
     91 ACGCGACTTCG 81
        :::::::::|||
    826 erSerSerSer 829

seq_name: sp_rodent:Q99K60

seq_documentation_block:
ID   Q99K60      PRELIMINARY;      PRT;   631 AA.
AC   Q99K60;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MUCIN 1, TRANSMEMBRANE.
GN   MUC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC005441; AAH05441.1; -.
DR   MGD; MGI:97231; Muc1.
DR   InterPro; IPR000082; SEA.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Transmembrane.
SQ   SEQUENCE   631 AA;  64690 MW;  B4236DF66655F4C3 CRC64;


alignment_scores:
           Quality:  122.00            Length:    467
             Ratio:  0.604              Gaps:     22
 Percent Similarity:  43.255   Percent Identity: 23.769

alignment_block:
 US-09-303-518D-131/rev x Q99K60   ..

 Align seg 1/1  to: Q99K60  from: 1  to: 631

    1337 TCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTA 1288
         |||   |||::::::|||:::|||
      38 SerSerSerLeuAlaSerThrThr..................... 45

    1287 TTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATT 1238
                     |||   :::|||:::      |||:::|||:::   :
      46 ............ThrProValHisSer......SerAsnSerAspProA 56

    1237 CCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCG 1188
         ::   :::|||   :::   :::   |||   ||||||:::   :::|||
      56 laThrArgProProGlyAspSerThrSer...SerProValGlnSerSer 71

    1187 CGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGT 1138
            :::|||   :::   |||      :::      |||   :::   ||
      72 ThrSerSerProAlaThrArgAlaProGluAspSerThrSerThrAlaVa 88

    1137 GCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACT 1088
         |      ||||||       ||||||      ||||||   ||||||
      88 lLeuSerGlyThr......SerSerProAlaThrThrAlaProValAsnS 103

    1087 TGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAG 1038
              |||                                      :::
     103 erAlaSer.....................................Ser 106

    1037 TATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCC 988
            ::::::         ::::::::::|||   :::||||||   :::
     107 ProValAlaHisGlyAspThrSerSerProAlaThrSerLeuSerLysAs 123

     987 TTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCG 938
          |||:::   :::   :::            :::         |||   |

     123 pSerAsnSerSerProVal.........ValHisSerGlyThrSerSerA 137

     937 CGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTTG 888
         |||||::::::   ||||||   :::
     137 laProAlaThrThrAlaProValAsp..................... 145

     887 TCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAA 838
               ||||||:::||||||::|||          |||:::::|||
     146 ......SerThrSerSerProValValHisGlyGlyThrLeuSerProAl 160

     837 AACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAA 788
          |||      :::      |||      |||      |||      ::::
     160 aThr......SerProProGlyAspSerThrSerSerProAspHisSerS 175

     787 CCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCA 738
         ::|||   |||           |||:::|||   :::   |||
     175 erThrSerSer............ProAlaThr...ArgAlaProGluAsp 186

     737 ATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGG 688
            ||||||                |||:::::::::::   |||::
     187 SerThrSer.................ThrAlaValLeuSerGlyThrSe 197

     687 CTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGC 638
         :|||                              ||||||      :::|||
     197 rSer..........................ProAlaThrThrAlaProV 205

     637 CAAATTCATGTGTTTCGATATTGGCA....................... 612
           :::|||      |||   :::|||
     205 alAspSerThrSerSerProValAlaHisAspAspThrSerSerProAla 221

     611 ..............................GCATTTTCAGACGGCAC 595
                                       :::         |||||
     222 ThrSerProSerGluAspSerAlaSerSerProValValHisGlyGlyTh 238

     594 GTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGC 545
         ||||:::||||||:::                        |||      |||
     238 rSerSerProAlaThr....................SerProLeuArgA 248

     544 TCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATG 495
           :::|||:::   |||      :::|||::::::   :::   :::
     248 spSerThrSerSerProValHisSerSerAlaSerIleGlnAsnIleLys 264

     494 ACCGTAGGGTCGGCAGCCAGC............GGATTGGTGTCCATCGC 457
         |||::::::::::   ||||||           :::   :::|||:::::
     265 ThrThrSerAspLeuAlaSerThrProAspHisAsnGlyThrSerValTh 281

     456 ATTGACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGA 407
         :   |||:::   |||   |||||||||:::::
     281 rThrThrSerSerAlaLeuGlySerAlaThrSer................ 292

     406 ACGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGG 357
                                        ||||||
     293 ............................ProAsp............... 294

     356 CGCACTTTTTCGCTGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTC 307
                                     |||::::::::::|||      ::
     295 .......................HisSerGlyThrSerThrThrThrAs 303

     306 GAACTCGATTTCGTCGTTGCCTTCAACGGCAATCACGACTGACTGAAGTA 257
         ::::|||   |||:::|||   :::|||   :::   :::      |||
     303 nSerSerGluSerAspLeuAlaThrThrProValTyrSer......SerM 318

     256 CGCGCTTTTCGCCACGGTGAATA.........GCGGCGATTTTGCCTGAA 216
               ||||||        :::         :::|||||||:::|||:::
     318 etProPheSerThrThrLysValThrSerGlySerAlaIleIleProAsp 334

     216 ............................................... 216

     335 HisAsnGlySerSerValLeuProThrSerSerValLeuGlySerAlaTh 351

     215 ................GCCGGCGCAGTAAATACTACGCCCGGATTCTTTT 182
                         ::::::|||:::   |||||||||
     351 rSerLeuValTyrAsnThrSerAlaIleAlaThrThrPro.......... 364
```

```
     181 TGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATT 132
365 ..ValSerAsnGlyThrGlnPro.........SerValProSerGlnTyr 377

     131 TTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTC 82
378 ProValSerProThrMetAlaThrThrSerSerHisSerThrIleAlaSe 394

      81 G 81
394 r 394
```

seq_name: sp_human:Q9UF83

seq_documentation_block:

```
ID   Q9UF83      PRELIMINARY;      PRT;   580 AA.
AC   Q9UF83;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 61.1 KDA PROTEIN (FRAGMENT).
GN   DKFZP434C196.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Koehrer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL133561; CAB63715.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   580 AA;  61089 MW;  C46E82F050F06547 CRC64;
```

alignment_scores:

```
            Quality:  121.50            Length:    530
              Ratio:  0.557               Gaps:     23
 Percent Similarity:  41.132   Percent Identity: 23.019
```

alignment_block:

US-09-303-518D-131/rev x Q9UF83 ..

Align seg 1/1 to: Q9UF83 from: 1 to: 580

```
    1336 CCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTAT 1287
      29 ProSerArgAlaSerProThrArgMetProSerArgAlaSerLeuLysMe 45

    1286 TTGCCCGGGCAGACGAAGCTGCACAAAG.........CGAGGTCTTCTTC 1246
      45 tThrProPheArgAlaSerLeuThrLysMetGluSerThrAlaLeuLeuA 62

    1245 GTCCAATTCCA..................................... 1235
      62 rgThrLeuProArgAlaSerLeuMetArgThrProThrArgAlaSerLeu 78

    1234 .....AGCAA 1230
      79 MetArgThrProProArgAlaSerProThrArgLysProProArgAlaSe 95

    1229 CCCAAAG...CCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCAA 1183
      95 rProArgThrProSerArgAlaSerProThrArgArgLeuProArgAlaS 112

    1182 AAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGA 1133
     112 erPro......MetGlySerProHis......ArgAlaSerProMetArg 124

    1132 TCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAAG 1083
     125 ThrProPro.....................ArgAlaSerProThr..... 132

    1082 AGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTT 1033
     133 ...................GlyThrProSerThrAlaSerProThrGlyT 143

    1032 GTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTT 983
     143 hrProSerSerAlaSerProThrGlyThrProPro.............. 154

     982 CGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCT 933
     155 ......ArgAlaSerProThrGlyThrProProArgAlaTrpAlaThrAr 169

     932 TGTGCAATCGCACCGTTCAATA.........CCGAACCGGAAATCA.... 896
     169 gSerProSerThrAlaSerLeuThrArgThrProSerArgAlaSerLeuT 186

     895 .....CGCGGTTGT 887
     186 hrArgTrpProProArgAlaSerProThrArgThrProProArgGluSer 202

     886 CCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTC....GCACC 841
     203 ProArgMetSerHisArgAlaSerProThrArgThrProProArgAlaSe 219

     840 CAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGG 791
     219 rProThrArgArgProProArgAlaSerProThrArgThrProProArgG 236

     790 CAACCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATA 741
     236 luSerLeuArgThrSerHisArgAlaSerProThrArgMetProProArg 252

     740 GCAATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGAC 691
     253 AlaSerProThr............ArgArgProProArgAlaSerProTh 265

     690 TGGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGG. 642
     265 rGlySerProPro...ArgAlaSerProMetThrProProArgAlaSerP 281

     641 .....CCGCCAAATTCATGT 627
     281 roArgThrProProArgAlaSerProThrThrThrProSerArgAlaSer 297

     626 GTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTT 577
     298 LeuThrArgThrProSerTrpAlaSerProThrThrThrProSerArgAl 314

     576 ACACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGC 527
     314 aSerLeuMetLysMetGluSerThrValSerIleThrArgThrProProA 331

     526 GTTTGAAG.....................................TCT 516
     331 rgAlaSerProThrGlyThrProSerArgAlaSerProThrGlyThrPro 347

     515 TCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAGCGGATTGGT 466
     348 SerArgAlaSerLeu......ThrGlySerProSerArgAlaSerLeuTh 362

     465 GTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGA 416
     362 rGlyThrProSerArgAlaSerLeuIleGlyThrProSerArgAlaSerL 379

     415 TTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATC 366
     379 euIleGlyThrProSerArgAlaSerLeuThrGlyThrProProArgAla 395
```

```
    365 AGGTTGCGGCGCACTTTTTCGCTGCTCAATTTT..................  333
        |||      |||   |||      ::::::
    396 SerLeuThrGlyThrSerSerThrAlaSerLeuThrArgThrProSerAr  412

    332 .GCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTT  284
         ||||||   :::   |||      |||:::|||::::::      |||
    412 gAlaSerLeuThrArgThrGlnSerSerSerSerLeuThrArgThrProS  429

    283 CAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATA  234
        ||   |||                  ||||||      ||||||
    429 erMetAlaSerLeu............ThrArgThrProProArg......  439

    233 GCGGCGATTTTG.........CCTGAAGCCGGCGCAGTAAATACTACGCC  193
        |||::::::            |||:::|||:::   ::::::|||   ||
    440 AlaSerLeuThrArgThrProProArgAlaSerLeuThrArgThrProPr  456

    192 CGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCAC  143
        |        |||:::   :::                 ||||||   :::|
    456 oArgAlaSerLeuThrArgThrProProArgAlaSerLeuThrArgThrP  473

    142 CTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCG......  99
        |||||::::::   ::::::   |||   |||:::   :::|||  *
    473 roSerMetValSerLeuLys...ArgSerProSerArgAlaSerLeuThr  488

     98 ......CCAAGCAACGCGACT.....................TCGGT  79
              ||||||:::|||:::                       |||::
    489 ArgThrProSerArgAlaSerLeuThrMetThrProSerArgAlaSerLe  505

     78 AATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCCC  39
        :        ||||||      :::   :::|||   |||
    505 uThrArgThrProSerThrAlaSerLeuThrGlyThrPro  518
```

```
seq_name: sp_fungi:O94317

seq_documentation_block:
ID   O94317      PRELIMINARY;      PRT;   534 AA.
AC   O94317;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 53.4 KDA SERINE-RICH PROTEIN C215.13 IN CHROMOSOME II
DE   PRECURSOR.
GN   SPBC215.13.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- SIMILARITY: TO YEAST STA1.
DR   EMBL; AL033534; CAA22127.1; -.
KW   Hypothetical protein; Glycoprotein; Signal.
FT   SIGNAL        1     14       POTENTIAL.
FT   CHAIN        15    534       HYPOTHETICAL SERINE-RICH PROTEIN C215.13.
FT   CARBOHYD     31     31       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   534 AA;  53361 MW;  D62DC15C80A698E4 CRC64;


alignment_scores:
           Quality:  121.00            Length:    418
             Ratio:  0.605               Gaps:     14
 Percent Similarity:  47.847   Percent Identity: 20.574

alignment_block:
 US-09-303-518D-131/rev x O94317   ..

 Align seg 1/1  to: O94317  from: 1  to: 534
```

```
   1280 GGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCA 1231
        |||   |||      :::   :::   ::::::   :::::::|||:::::
     67 GlyGlyThrProThrTyrTyrGlyTyrAlaThrProThrSerSerSerGl  83

   1230 ACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACG...ATTAAATCGCGCA 1184
        :|||:::      :::   |||::::::|||:::      :::|||   :
     83 uProSerIlePheSerGluSerAlaThrProSerGluThrAsnSerTyrS  100

   1183 AAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCG 1134
        ::|||   |||:::      |||:::      |||   |||   :::|||
    100 erSerProValSerSerTyrSerAspProAlaThr...SerGlnLeuPro  115

   1133 ATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAA 1084
           :::|||         ||||||      :::
    116 SerSerThrSerPhePheSerProThrSerSer.................  126

   1083 GAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATT 1034
                                                  ||||||
    127 .....................................GluTyrT  129

   1033 TGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCT 984
                     ::::::      :::||||||||||||   ||||||
    129 hrPro.........SerSerThrGluSerSerSerLeuLeuAspProSer  142

    983 TCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCC 934
        |||::::::::
    143 SerValSerSer.....................................  146

    933 TTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTTGTCCG 884
           ||||||   |||   :::|||:::   |||::::::      |||:
    147 ....AlaIleLeuProSerSerThrSerValGluValSerIleSerSerS  162

    883 CGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACG 834
        ::|||   :::|||               |||   ::::::   :::|||
    162 erSerLeuSerSerSer............AspProLeuThrSerSerThr  174

    833 GTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCC......AAGGC 790
              :::      :::      :::          |||      :::::
    175 PheSerSerLeuSerSerSerThrSerSerSerGlnProSerValSerSe  191

    789 AACCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAG 740
        :|||:::   |||:::|||
    191 rThrSerSerSerThrPhe.............................  197

    739 CAATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACT 690

    197 .....................................................  197

    689 GGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCC 640
                                                :::::::::::||
    198 .................................SerSerAlaAlaPr  202

    639 GCCAAATTCATGTGTTTCGATATTGGCAGCATTTTCA......GACGGCA 596
        |   ::::::      |||   |||::::::   |||      :::::::
    202 oThrSerThrSerSerSerTyrLeuSerSerSerSerValValSerSerS  219

    595 CGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGG 546
        ::|||:::|||::::::      :::      |||   |||
    219 erSerSerProSerSerSerSerSerSerThrLeuThrSerSerSerLeu  235

    545 CTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGAT 496
           :::::::::   |||      :::||||||:::::::|||
    236 SerThrSerSerIleProSerThrSerSerSerSerSerSerThrSerSe  252

    495 GACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCATTGACGAAGA 446
         ::::::::::|||:::::::|||                  ::::::
    252 rSerLeuSerSerSerSerSer.....................SerSerT  262

    445 TGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGACGGGTG 396
         |||::::::|||:::|||::::::::::|||:::   ::::::
    262 hrAlaSerSerSerSerSerSerSerSerIleIleSerSerSer......  276
```

```
     395 CGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTC 346
         |||:::      :::|||                        |||:::||
     277 SerSerSerSerSerSerPro......ThrSerThrSerSerThrIleSe 291

     345 GCTGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTT 296
         |      :::   :::|||:::|||   |||      |||::::::::||||
     291 rSerSerSerSerSerSerSerSerProThrSerThrSerSerThrIleS 308

     295 CGTCGTTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCG 246
         |||||      |||:::                  |||:::   ||||||
     308 erSerSerSerSerSer.................SerSerSerPheSer 318

     245 CCACGGTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTAC 196
                     ::::::   :::   ::::::::::::   ::::::::
     319 SerThrLeuSerSerSerSerMetSerSerSerSerSerPheSerSerSe 335

     195 GCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTT 146
         :|||           ||||||:::|||||             :::::|
     335 rPro.........ThrSerSerSerSerThrIleSerSerSerSerSerS 349

     145 CACCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCA 96
         ||||||||          :::::::::          :::   :::|||
     349 erProSerSerSerSerPheSerSerThrThrSerSerSerLysSerSer 365

      95 AGCAACGCGACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGG 46
         |||:::   :::|||:::::::::::::   |||      :::   |||
     366 SerSerPheSerSerThrValSerSerSerSerSerThrSerSerSerTh 382

      45 TCTG 42
         |||
     382 rLeu 383

seq_name: sp_human:Q9UQ39

seq_documentation_block:
ID   Q9UQ39      PRELIMINARY;      PRT;   956 AA.
AC   Q9UQ39;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA BINDING PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016088; BAA83714.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER       1      1
FT   NON_TER     956    956
SQ   SEQUENCE   956 AA;  106462 MW;  C9FEAD07B49F5011 CRC64;


alignment_scores:
           Quality:  118.50            Length:    398
             Ratio:  0.598               Gaps:     19
 Percent Similarity:  49.749   Percent Identity: 25.126

alignment_block:
 US-09-303-518D-131 x Q9UQ39   ..

 Align seg 1/1  to: Q9UQ39  from: 1  to: 956

      18 AGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACG 67
         ||||||:::||||||:::||||||:::      ||||||
     518 ArgSerArgSerAlaArgArgGlyArg.........SerHisSerArgSe 531

      68 GCCC.........GGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATAT 108

         |||         |||:::   |||||||||:::      ||||||
     531 rProAlaThrArgGlyArgSerArgSerArgThrProAlaArgArg.... 546

     109 GTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAA 158
                                          ||||||   ||||||::::::
     547 ............................GlyArg...SerArgSerArg 552

     159 AGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTG 208
         |||:::         |||:::::::::   :::|||:::
     553 ThrProAla.........ArgArgArgSerArgSerArgThr........ 563

     209 CGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTA 258
                 |||:::::::|||      :::|||   |||::::::|||
     564 .ProThrArgArgArgSerArgSerArgThrProAlaArgArgGlyArg. 579

     259 CTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGA 308
               |||||||:::            :::|||||||:::|||:::|||
     580 ........SerArgSerArgThrProAlaArgArgArgSerArgThrArg 593

     309 ACG...CTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGC 355
         :::   :::|||      |||   ::::::         |||:::|||::
     594 SerProValArgArgArgSerArgSerArgSerProAlaArgArgSerGl 610

     356 GCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCG 405
         :                                                |||
     610 y.........................................ArgSerA 613

     406 TTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAA 455
           ::::::::::|||:::||||||   |||:::   |||
     613 rgSerArgThrProAlaArgArgGlyArgSer...ArgSerArgThrPro 628

     456 TGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAG 505
         :::         :::|||   :::      |||      :::      ||
     629 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr 645

     506 AAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACC 555
         |||||||   |||      :::|||            :::      :::|
     645 gSerArgSerArgThrProAlaArg...........ArgGlyArgSerA 658

     556 GAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGA 605
         ||:::   :::|||           ||||||   |||      |||
     658 rgSerArgThrProArgArgGlyArgSerArgSerArgSerLeuValArg 674

     606 AAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCG 655
         :::                                |||      |||
     675 Arg.........................GlyArgSerHisSerArgTh 682

     656 GCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAA 705
         :::   :::::::::|||       ::::::|||::::::   :
     682 rProGlnArgArgGlyArgSerGlySerSerSerGluArgLysAsnLysS 699

     706 ACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTT 755
         ::|||:::::::::::   ||||||   :::           :::
     699 erArgThrSerGlnArgArgSerArgSerAsnSerSerProGluMetLys 715

     756 CGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGC 805
         ::::::|||   |||:::   |||:::|||:::   |||:::   |||
     716 LysSerArgIleSerSerArgArgSerArgSer...LeuSerSerProAr 731

     806 AAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCT 855
          |||:::      :::   |||:::   |||   :::   :::|||   |
     731 gSerLysAlaLysSerArgLeuSerLeuArgArgSerLeuSerGlySerS 748

     856 CAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGG 905
         ||        ::::::   :::   :::      ||||||      |||
     748 erProCysProLysGlnLysSer...GlnThrProProArgArgSerArg 763

     906 TTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGAC 955
           |||                                        :::::
     764 SerGly.....................................SerSe 767

     956 GCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTG 1005
         :   |||::::::      |||         ||| ||||||  :::|||
```

```
     767 rGlnProLysAlaLysSerArgThrProProArg.ArgSer.ArgSerSe 783

    1006 TTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCAC 1055
          |||:::         ::::::::::::::::||||||||||||
     783 rSerSerProProProLysGlnLysSerLysThrProSerArg..... 797

    1056 TCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCA 1105
                        :::::::::||||||||||||   :::|||:::
     798 ................GlnSerHisSerSerSerSerProHisProLys 808

    1106 ACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATG 1155
          :::   :::   :::|||      ::::::::::::   |||      ::
     809 ValLysSerGlyThrProProArgGlnGlySerIleThrSerProGlnAl 825

    1156 CCGTTGGACATCCTGCCTACCT............TGCTTT 1183
          ::::      |||      |||            ||||||
     825 aAsnGluGlnSerValThrProGlnArgArgSerCysPhe 838

seq_name: sp_human:Q9UQ40

seq_documentation_block:
ID   Q9UQ40      PRELIMINARY;      PRT;  1262 AA.
AC   Q9UQ40;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RNA BINDING PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016087; BAA83713.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER      1      1
FT   NON_TER    1262   1262
SQ   SEQUENCE   1262 AA;  138003 MW;  75E76BD73FA14295 CRC64;


alignment_scores:
            Quality:  118.50              Length:    398
              Ratio:  0.598                Gaps:     19
 Percent Similarity:  49.749   Percent Identity: 25.126

alignment_block:
 US-09-303-518D-131 x Q9UQ40   ..

 Align seg 1/1  to: Q9UQ40  from: 1  to: 1262

      18 AGGTCTAAATCTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACG 67
         ||||||:::|||||||:::||||||:::          ||||||
     343 ArgSerArgSerAlaArgArgGlyArg.........SerHisSerArgSe 356

      68 GCCC.........GGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATAT 108
          |||          |||:::   |||||||||:::      ||||||
     356 rProAlaThrArgGlyArgSerArgSerArgThrProAlaArgArg.... 371

     109 GTCGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAA 158
                                        ||||||   ||||||::::::
     372 ............................GlyArg...SerArgSerArg 377

     159 AGGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTG 208
             |||:::         |||:::::::::   :::|||:::
     378 ThrProAla.........ArgArgArgSerArgSerArgThr........ 388

     209 CGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTA 258
                 |||::::::|||      :::|||   |||::::::|||
     389 .ProThrArgArgArgSerArgSerArgThrProAlaArgArgGlyArg. 404

     259 CTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGA 308
              ||||||:::              :::|||||||:::|||:::|||
     405 ........SerArgSerArgThrProAlaArgArgArgSerArgThrArg 418

     309 ACG...CTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGC 355
         :::   :::|||      |||   ::::::         |||:::|||::
     419 SerProValArgArgArgSerArgSerArgSerProAlaArgArgSerGl 435

     356 GCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCG 405
         :                                                |||
     435 y.......................................ArgSerA 438

     406 TTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAA 455
           :::::::::|||:::||||||   |||:::   |||
     438 rgSerArgThrProAlaArgArgGlyArgSer...ArgSerArgThrPro 453

     456 TGCGATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAG 505
         :::          :::|||   :::      |||      :::      ||
     454 AlaArgArgGlyArgSerArgSerArgThrProAlaArgArgSerGlyAr 470

     506 AAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACC 555
         |||||||   |||      :::|||            :::      :::|
     470 gSerArgSerArgThrProAlaArg...........ArgGlyArgSerA 483

     556 GAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGA 605
         ||:::   :::|||            ||||||   |||      |||
     483 rgSerArgThrProArgArgGlyArgSerArgSerArgSerLeuValArg 499

     606 AAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCG 655
         :::                              |||      |||
     500 Arg.........................GlyArgSerHisSerArgTh 507

     656 GCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAA 705
             :::   ::::::::::|||        ::::::|||::::::   :
     507 rProGlnArgArgGlyArgSerGlySerSerSerGluArgLysAsnLysS 524

     706 ACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTT 755
         ::|||:::::::::::   ||||||   ::            ::
     524 erArgThrSerGlnArgArgSerArgSerAsnSerSerProGluMetLys 540

     756 CGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGC 805
         ::::::|||   |||:::   |||:::|||:::   |||:::   |||
     541 LysSerArgIleSerSerArgArgSerArgSer...LeuSerSerProAr 556

     806 AAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCT 855
          |||:::      :::   |||:::   |||   :::   :::|||   |
     556 gSerLysAlaLysSerArgLeuSerLeuArgArgSerLeuSerGlySerS 573

     856 CAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGG 905
         ||         ::::::   :::   :::      ||||||      |||
     573 erProCysProLysGlnLysSer...GlnThrProProArgArgSerArg 588

     906 TTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGAC 955
            |||                                         :::::
     589 SerGly.....................................SerSe 592

     956 GCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTG 1005
         :   |||::::::      |||         ||| ||||||  :::|||
     592 rGlnProLysAlaLysSerArgThrProProArg.ArgSer.ArgSerSe 608

    1006 TTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCAC 1055
          |||:::         :::::::::::::::||||||||||||
     608 rSerSerProProProLysGlnLysSerLysThrProSerArg..... 622

    1056 TCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCA 1105
                        :::::::::||||||||||||   :::|||:::
     623 ................GlnSerHisSerSerSerSerProHisProLys 633

    1106 ACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATG 1155
          :::   :::   :::|||      ::::::::::::   |||      ::
     634 ValLysSerGlyThrProProArgGlnGlySerIleThrSerProGlnAl 650
```

```
   1156 CCGTTGGACATCCTGCCTACCT.............TGCTTT 1183
          ::::     |||      |||             ||||||
    650 aAsnGluGlnSerValThrProGlnArgArgSerCysPhe 663

seq_name: sp_plant:Q9SUV0

seq_documentation_block:
ID   Q9SUV0      PRELIMINARY;      PRT;   857 AA.
AC   Q9SUV0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 96.9 KDA PROTEIN.
GN   F8B4.120 OR AT4G32420.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bevan M., Terryn N., Ardiles W., Buysshaert C., Dasseville R.,
RA   De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA   Villaroel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W.,
RA   Mayer K.F.X., Lemcke K., Schueller C.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Terryn N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,
RA   De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
RA   Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL034567; CAA22569.1; -.
DR   EMBL; AL161581; CAB79959.1; -.
DR   HSSP; P05092; 2CPL.
DR   InterPro; IPR002130; CSA_PPIase.
DR   Pfam; PF00160; pro_isomerase; 2.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   857 AA;  96854 MW;  17E4FA842D58D972 CRC64;


alignment_scores:
           Quality:  118.00              Length:    438
             Ratio:  0.621                Gaps:      21
 Percent Similarity:  43.379   Percent Identity: 24.429

alignment_block:
 US-09-303-518D-131 x Q9SUV0   ..

 Align seg 1/1  to: Q9SUV0  from: 1  to: 857

     111 CGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAA..... 155
         |||::::::|||   :::|||::::::|||||      |||:::
     448 ArgGluSerProGlySerGluGluLysGlyArgHisValArgArgSerPr 464

     156 ...............AAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAAT 189
                         :::   |||      |||   ::::::::::
     464 oThrLysSerValSerArgSerProValArgValLysLysGluArgAspI 481

     190 CCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCA 239
           :::||||||                        ::::::      |||
     481 leSerArgSerPro.............SerLysSerLeuSerArgSer 492

     240 CCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGAT...TGCCGTTGAAG 286
         |||   |||:::   :::      :::|||||:::      |||   ||
     493 ProLeuArgSerProLysArgValIleSerArgSerProValArgGlyAr 509

     287 GCAACGACGAAATCGAGTTCGAACG...CTACGTACCTGAAGCGCTGGCA 333
         |      |||:::      |||:::   :::|||:::   |||   |||:
     509 gIleAlaArgSerProSerArgSerProValArgSerAlaSerArgGlyS 526

     334 AAATTGAGCAG..................................... 344
         :::::   :::
     526 erLeuGlyArgGlyProLeuArgArgSerSerArgArgSerProSerArg 542

     345 .........CGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGA 385
                  |||:::||| ||||||:::|||   :::|||   :::
     543 SerProValArgSerSer.ArgArgSerLeuSerArgSerProIleGlnL 559

     386 CTGCGCTTCGC.................ACCCGTCCGTTCAGCAAAATC 417
           :::   |||                 ||||||    |||   :::||
     559 euSerArgArgSerLeuSerArgSerProThrArg.LeuSerArgArgSe 575

     418 CCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACAC 467
         ||||         |||   |||||||||   :::|||:::      :::|
     575 rLeuSerArgSerProIleArgSerProArgLysSerValSerArgSerP 592

     468 CAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAG 517
         ||:::|||         :::   :::|||   |||::::::|||   :::
     592 roValArg......SerSerArgLysSerValSerArgSerProValArg 606

     518 ACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATC 567
         :::|||::: |||   ::::::   :::|||
     607 SerSerArg.ArgArgIleSerArgSerPro..................... 616

     568 CATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAA 617
               |||   |||||||||:::
     617 .....ValArgSerSerArgLysSer........................ 623

     618 TATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCA 667

     624 .............................................Va 624

     668 CGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACC 717
          ::::::|||      |||   ||||||||:::   :::|||
     624 lSerArgSerProIleArgLeuSerArgArgSerIleSerArgSerProI 641

     718 ATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGG... 764
           :::|||||||||||||:::      |||:::   ||||||:::|||
     641 leArgLeuSerArgArgSerIleSerArgSerProValArgGlyArgArg 657

     765 .....................CCGTCTGAATACCGAGCG........... 782
                              |||:::::   |||:::
     658 ArgIleSerArgSerProValProAlaArgArgArgSerValArgProAr 674

     783 ...................CGTGGTTGCCTTGGGCGGCCTGCAAGT.... 809
                           |||      |||:::|||   ||||||
     674 gSerProProProAspArgArgArgSerLeuSerArgSerAlaSerProA 691

     810 ..........CAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAA 848
                    ::::::      ::::::   ||||||
     691 snGlyArgIleArgArgGlyArgGlyPheSerGlnArgPhe......... 704

     849 GGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCG..... 893
                ::|||   ||||||              :::|||
     705 .........SerTyrAlaArgArgTyrArgThrSerProSerProAspAr 718

     894 .CGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGG....... 935
              :::|||||||:::   :::|||   |||      |||
     718 gSerProTyrArgPheSer...AspArgSerAspArgAspArgPheArgS 734

     936 ............................CGCGCATGATTATTTGGGACG 956
                                     |||:::   |||   |||
     734 erArgArgArgPheSerProSerArgPheArgSerProLeuArgGlyArg 750

     957 CTACCACAATCAGATTTCCGTTATCGAAGAAGGCC.............. 991
```

```
          |||   |||   :::|||||||||||||||||
     751 ThrProProSerMetLeuArgTyrArgArgArgSerArgSerValSerPr 767

     991 ..................................................... 991

     767 oGlyLeuCysTyrArgAsnArgArgTyrSerArgSerProIleArgSerA 784

     992 ................................GCAGCAAAGAGCTGT 1006
                                         |||:::   |||
     784 rgSerProProTyrArgLysArgArgSerProSerAlaSerHisSerLeu 800

    1007 TCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACT 1056
         |||   :::      |||||||||:::::::::   |||::::::|||::
     801 SerProSerArgSerArgSerArgSerLysSerTyrSerLysSerProIl 817

    1057 C.........TCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGAC 1097
         :         :::|||   |||   :::   :::   |||::::::|||:
     817 eGlyThrGlyLysAlaArgSerValSerArgSerProSerLysAlaArgS 834

    1098 AGCCGTCA 1105
         ::||||||
     834 erProSer 836

seq_name: sp_mammal:Q29071

seq_documentation_block:
ID   Q29071      PRELIMINARY;      PRT;   528 AA.
AC   Q29071;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GASTRIC MUCIN (FRAGMENT).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX   MEDLINE=95275264; PubMed=7755593;
RA   Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA   LaMont J.T.;
RT   "Isolation and characterization of cDNA clones encoding pig gastric
RT   mucin.";
RL   Biochem. J. 308:89-96(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX   MEDLINE=94102478; PubMed=7506218;
RA   Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA   LaMont J.T.;
RT   "Pig gastric mucin: isolation and characterization of a cDNA clone
RT   with a novel tandem repeat.";
RL   Gastroenterology 106:200-200(1994).
DR   EMBL; U10281; AAC48526.1; -.
FT   NON_TER       1      1
FT   NON_TER     528    528
SQ   SEQUENCE   528 AA;  49907 MW;  0BF0F6879203B2EA CRC64;


alignment_scores:
            Quality:  116.00             Length:    441
              Ratio:  0.518                Gaps:     19
 Percent Similarity:  50.794   Percent Identity: 22.222

alignment_block:
 US-09-303-518D-131/rev x Q29071   ..

 Align seg 1/1  to: Q29071  from: 1  to: 528

    1343 CAGCCTTCCTTCTCAATG.........GTTTCCAGCACTTTGCGCAACAG 1303
         |||||||||   |||            :::||||||||::
     149 GlnProSerSerSerSerSerProProIleSerSerThrVal.......: 162

    1302 CGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGT 1253
               |||      |||:::::::::::::      :::|||   :
     163 .........SerValGlnProSerSerSerSer......SerAlaProT 174

    1252 CTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCG 1203
         :::::|||:::::::|||   ||||||::::::         |||   |||
     174 hrThrSerAlaThrSerValGlnProSerSer.........SerSerSer 187

    1202 CCGACGATTAAATCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCAT 1153
         |||   |||:::|||      |||               |||::::::::
     188 ProProIleSerSerThrValSerValGlnThrSerSerSerSerSerVa 204

    1152 T............ACGCGCTCATAAGTGCCGATCGGTACCATGGCGCGGT 1115
         :            |||   |||      |||      :::   :::   |
     204 lProThrThrSerThrThrSerValGlnPro......SerSerSerSerS 219

    1114 CGCCGCCGTTGACGGCTGTCGTGAACTTGAAGAGTTTGTTTTTTAGGAAA 1065
         ||   |||   |||:::::::::::::::::::|||            :::
     219 erValProThrThrSerAlaThrSerValArgSerSerSerSerSerSer 235

    1064 TGGCCGAGAGTGGTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAAC 1015
            |||                        :::   :::      :::::
     236 ThrPro.....................IleProSerThrThrSerVa 244

    1014 CCAGCCGAACAGCTCTTTGCTGCGGCCT...TCTTCGATAACGGAAATCT 968
         :||||||::||||||         |||   ::::|||   |||::::::
     244 lGlnProSerSerSerSerSerAlaProThrThrSerAlaThrSerVal. 260

     967 GATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCG 918
                 :::|||:::   |||   :::            :::
     261 ..........GlnProSerSerSerSerSerThrProIleProSerThr 273

     917 TTCAATACCGAACCGGAAATCACGCGGTTGTCCGCGTCAACCAATTCGCC 868
         :::::::::|||       |||:::|||::::::::::||
     274 ThrSerValGlnPro.............SerSerSerSerSerAlaPr 285

     867 GGCGGTAAGTTGAGACACCTTCGCACCCAAAACGGTACGCAAGAGGCGCG 818
         |::::::|||      :::      |||::::::
     285 oThrThrSerAlaThrSerValGlnProSerSer............... 296

     817 GTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGA 768
               :::      ||||||   ::::::|||   ||||||
     297 ........SerSerSerProProIleSerSerThrIleSerValGlnPro 310

     767 CGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGAT 718
               :::::::::         :::   ||||||      :::::
     311 SerSerSerSerSerSerProThrThrSerThrThrSer......ValGl 325

     717 GGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCG 668
         :      :::         |||||||||   |||            ::::
     325 nProSerSerSerGlySerAlaProThrThrSer............AlaT 338

     667 TGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATA 618
         ::   ::::::|||::::::   :::   |||            :::|||
     338 hrSerValGlnProSerSerSerSerSerProProIleSerSerThrIle 354

     617 TTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATG 568
            :::      |||:::::::::::|||:::|||::::::      |||
     355 SerValGlnProSerSerSerSerSerSerProThrThrSerThrThrSe 371

     567 GATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGT 518
           :::                          :::|||      ::::
     371 rVal............................GlnProSerSerSerG 378

     517 CTTCGGCGGCTTCTTTGATGATGACC......GTAGGGTCGGCAGCCAGC 474
         ::||||||   :::         |||            |||:::::|||
     378 lySerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSer 394

     473 GGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTAC 424
         ::::::   :::   :::   |||::::::   ::::::|||:::|||::
     395 SerValProThrThrSerAlaThrSerValArgSerSerSerSerSerSe 411
```

```
     423 GGCAGGGATTTTGCTGAACGGACGGGTGCGA.............AGCGCAG 386
     411 rThrProIleProThrThrThrSerValGlnProSerSerSerSerServ 428

     385 TCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAAT 336
     428 alProThrThrSerAlaThrSerValGlnThrSerSerSerSerSerThr 444

     335 TTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCC 286
     445 ProIleProSerThrThrSerValGlnProSerSerSerSerSerAlaPr 461

     285 TTCAACGGCAATCACGACT...GACTGAAGTACGCGCTTTTCGCCACGGT 239
     461 oThrThrSerAlaThrSerValGlnProSerSerSerSerSerPro.... 476

     238 GAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGA 189
     477 .................ProIleSerSerThrIleSerValGlnPro... 486

     188 TTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTC 139
     487 .........SerSerSerSerSerSerProThrThrSerThrThrSerVa 500

     138 CTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACG 89
     500 lGlnProSerSerSerGlySerAlaProThrThrSerAlaThrSerValG 517

      88 CGACTTCGGTAATGGCCGGGCCG 66
     517 lnProSerSerSerSerSerPro 524

seq_name: sp_plant:Q9LWY3

seq_documentation_block:
ID   Q9LWY3      PRELIMINARY;      PRT;   739 AA.
AC   Q9LWY3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
RT   clone:P0644B06.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001129; BAA90631.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   739 AA;  82154 MW;  5DC56CAF5DD1CA88 CRC64;


alignment_scores:
            Quality:  116.50              Length:    436
              Ratio:  0.626                Gaps:     28
 Percent Similarity:  42.661   Percent Identity: 25.459

alignment_block:
 US-09-303-518D-131 x Q9LWY3   ..

 Align seg 1/1  to: Q9LWY3  from: 1  to: 739

      78 TACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGA 127
      10 HisArgAlaThrGlyGlyTrpGln.......................... 17

     128 TGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAA 177

      18 .GluGlyGlyGlyArg............GlnArgProThrAla......A 28

     178 GACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAAT 227
      28 rgArgGluGlyGlyGlyArgArg..................ArgAlaArg 38

     228 CGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTG 277
      39 ArgGluGlyGlyGlyTrpArgArgPro..................... 47

     278 CCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCG 327
      48 ............................ThrSerArgArgAspAspS 54

     328 CTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGG 377
      54 erGly.........GlnArgArgAlaArgProGlyGlyValThrLeuGln 67

     378 CTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAG 427
      68 IleLysArgLysProAsnGlyArgValValLeuSerIleProSerTyrTh 84

     428 ATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTG 477
      84 rThrThrAlaGlnProHis..............HisHisThrProAlaP 96

     478 GCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGA...AGACTTCAA 524
      96 roCysArgLeuValGlyArgHisGlnArgArgArgArgGlyThrLeuAsp 112

     525 ACGCGGCCTGTTGGTATTGAGCCGCCT.........GACCGAACGTAAAA 565
     113 ArgProProSerSerLeuLeuProProArgAlaIleAsp.SerArgArgL 129

     566 TCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCC 615
     129 euHis.....................AspHisProSer...AsnProAsp 137

     616 AATATCGAAACACATGAATTT............GGCGGCCCGCATCCTGC 653
     138 AsnIleProLeuLeuProLeuProThrLeuProSerProProHisPro.. 153

     654 CGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATA 703
     154 ..........GlyThrLeuIleArg..............ArgProSerA 162

     704 AAACCGTGTGGACCATC............AATTATCAAGACGTGATTGCT 741
     162 laGluValTrpValValThrSerAlaGlyGluGluGluGluValIleGly 178

     742 ATCGGACGTTTGTTCGTAACAGGCCGTCTGAA....TACCGAGCGCGTGG 787
     179 ValGlyGluAspAlaHisGlyAlaArgIleMetLysHisIleGlnArgGl 195

     788 TTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTT 837
     195 yCysLeuLysGluAlaAlaSerGlyAspAlaAlaValTrpArgCysAlaA 212

     838 TTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGT......... 878
     212 spGly.....................ArgArgArgSerGlyAsnGlyAsn 221

     879 ...TGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAA...CGGTG 922
     222 AspGluArgArgGlnCysSerAspArgArgTyrArgValGluValGluVa 238

     923 CGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATT 972
     238 lGluCysLeu.....................SerLeuProLeuArgGluA 248

     973 TCCGTTATCGAAGA............AGGCCGCAGCAAAGAGCTGTTCGG 1010
```

```
 248 laGluAlaHisArgAlaGluGluAlaArgProArgArgArgHisArgArg 264

1011 ..................CTGGGTTGCGCCGCAGCCGGACAAATACTCCA 1042
                       |||   ||||||   :::|||
 265 GluGluAlaLeuProTyrLeuCysCysAlaAspValGly........... 277

1043 TCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAA...G 1089
        |||::::::::::|||      |||:::         |||:::
 278 ....AlaGluArgLysArgGlnGluProArgAlaProArgLeuSerProP 293

1090 TTCACGACAGCCGTCAACGGCGGCGACCGCGC.................. 1121
       |||   |||   :::   ::::::||||||
 293 roHisArgSerGlySerAlaLysAsnProArgSerValArgLeuSerPro 309

1122 ..................CATGGTACCGATCGGCACTTATGAGCGCGTAA 1153
                        ||||||:::      :::      |||
 310 ProLeuLeuAlaSerLeuPheGlyThrAsnGlyAlaValProArgArgPr 326

1154 TGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGC 1203
               :::|||:::              |||::::::|||||||
 326 oProArgProArgProValProArgArgGlyArgArgLeuGlyArgArgA 343

1204 GATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGA 1253
     ||   ||||||                 ||||||:::
 343 rg...ArgGln.................LeuGlyLeuProAlaSerAla 352

1254 CCTCGC 1259
     ||||||
 353 ProArg 354
```

seq_name: sp_invertebrate:Q23587

seq_documentation_block:

```
ID   Q23587      PRELIMINARY;      PRT;  3507 AA.
AC   Q23587;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ZK783.1 PROTEIN.
GN   ZK783.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Favello A., Vaudin M.;
RT   "The sequence of C. elegans cosmid ZK783.";
RL   Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U13646; AAC24418.1; -.
DR   HSSP; P00736; 1APQ.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR001507; zona_pellucida.
DR   Pfam; PF00008; EGF; 15.
DR   SMART; SM00179; EGF_CA; 12.
DR   SMART; SM00001; EGF_like; 16.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 9.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 22.
DR   PROSITE; PS01187; EGF_CA; 13.
DR   PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW   Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ   SEQUENCE   3507 AA;  365315 MW;  154F0B687874D9DF CRC64;
```

alignment_scores:

```
            Quality:  115.00              Length:    415
              Ratio:  0.548                 Gaps:     20
 Percent Similarity:  50.602   Percent Identity: 22.410
```

alignment_block:

US-09-303-518D-131/rev x Q23587 ..

Align seg 1/1 to: Q23587 from: 1 to: 3507

```
1283 CCCGGGCAGACGAAGCTGCAC......AAAGCGAGGTCTTCTTCGTCCAA 1240
     |||:::::::   |||::::::      :::   :::|||||||||||||
1945 ProAsnArgGluLysIleGluIleAspGluGluAsnSerSerSerSerAs 1961

1239 TTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAAT 1190
     ||||   |||   |||                    ||||||   |||:
1961 nSerGlyGlnGluLys.....................ProThrThrLysG 1971

1189 CGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAA 1140
     ::      |||::::::::::      |||:::      |||   :::
1971 lyIleValSerSerThrSerAlaThrSerSerGluSerThrThrAlaGlu 1987

1139 GTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAA 1090
           :::   |||                 ::::::::::::::::
1988 ProHisValThrThr..............SerIleSerSerThrThrSe 1999

1089 CTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGG 1040
     :   |||::::::         |||   |||:::   |||   :::   :
1999 rThrLysAspMetThrSerSerLysSerProGluAsnValThrMetSerS 2016

1039 AGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGG 990
     ::                 ::::::::::::   :::|||:::
2016 erGluSerProGluValSerThrSerSerSerLysSerThrThrAlaSer 2032

 989 CCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATG 940
       ::::::::::::::                 |||:::   |||
2033 GluThrThrValSerSer..............ThrProSerGluSerSe 2044

 939 C.........GCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAA 899
               ||||||   :::   :::|||   :::|||         |
2044 rSerSerGluAlaProLeuThrSerSerProAlaThrThrThrGluValI 2061

 898 TCACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACC 849
     ||||||:::   |||:::::::::::::::|||           ::::::
2061 leThrGluSerSerValLysSerThrThrProLysGlu......GluSer 2075

 848 TTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCC 799
       :::      ||||||:::               |||:::   :::
2076 SerSerGluIleThrValLys..............LeuSerSerLysSe 2087

 798 GCCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGTTACG...AACA 752
      |||::::::|||      ||||||         |||   |||   :::
2087 rProGluValThrGluSerSerValLysSerSerProSerThrProSerT 2104
```

```
 751 AACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACACGGTTTTA 702
2104 hrThrSerGlnSerValThrSer.................ThrVal... 2113

 701 TTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGC 652

2113 ..................................................  2113

 651 AGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGCATTTTCAG 602
2114 .........ProGluThrSerLysSerThrValLeuSerSerGluAlaP 2127

 601 ACGGCACGTCT...GCGCCTGCTGCTTTACACACATGGATTTTACGTTCG 555
2127 roValThrSerThrSerProThrGluValHisThr.............. 2138

 554 GTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTC 505
2139 .........SerSerGluThrLysProSerLeuSerAlaSerSer..... 2150

 504 TTTGATGATGACCGTAGGGTCG.................GCAGCCAGCG 473
2151 .........ThrThrGlyAspThrAsnSerThrThrProSerThrSerS 2164

 472 GATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGC...TCGGCATCT 426
2164 erLeuAlaSerValLysSerThrSerAlaProGluGlyThrSerAlaSer 2180

 425 ACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCC 376
2181 ValAlaProValLysLeuSerSer...LeuSerProAspValSerGlnPr 2196

 375 TGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCG 326
2196 oSer............ThrLysThrPheAsp............AlaThrG 2205

 325 CTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGGCA 276
2205 luSerSerThrValGlnAlaSerGluThrSerSerGlyThrSerValLys 2221

 275 ATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCGAT 226
2222 SerThrSerGluProGluSerHisValThrLysLeuSerIleThrSerSe 2238

 225 TTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTTGTCTT 176
2238 rAsnProSerSerSerValProValThrSerPro............LysS 2251

 175 CAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATTTTCATC 126
2251 erThrProThrValProGluSerThrGluGlnProThrSerThrThrPro 2267

 125 GAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCG 81
2268 SerGlyGlnSerLeuThr.........ProMetAsnSerAsnSer 2279
```

seq_name: sp_invertebrate:Q23804

seq_documentation_block:

```
ID   Q23804      PRELIMINARY;      PRT;   749 AA.
AC   Q23804;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   SPID PRECURSOR (FRAGMENT).
GN   BR2.1.
OS   Chironomus tentans (Midge).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Chironomoidea; Chironomidae; Chironominae; Chironomus.
OX   NCBI_TaxID=7153;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92262483; PubMed=1584794;
RA   Wieslander L., Paulsson G.;
RT   "Sequence organization of the Balbiani ring 2.1 gene in Chironomus
RT   tentans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:4578-4582(1992).
DR   EMBL; M89909; AAA28264.1; -.
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   NON_TER     749    749
SQ   SEQUENCE   749 AA;  83564 MW;  606367E2800C45BD CRC64;
```

alignment_scores:

```
           Quality:  114.50              Length:    476
             Ratio:  0.500                 Gaps:     26
 Percent Similarity:  48.109   Percent Identity: 23.109
```

alignment_block:

US-09-303-518D-131 x Q23804 ..

Align seg 1/1 to: Q23804 from: 1 to: 749

```
  12 CAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGAC.CGGAGCAAGTCATT 60
317 GluLysArgSerLysSerGlySerArgProGluLysArgSerLysSerGl 333

  61 TATGACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGT 110
333 ySerArgProGluLysArgSerLysSerGlySerArgProGluLysArgS 350

 111 CGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAG 160
350 erGluSerGlyProArgProGluLysArgSerLysSerGlySerArgPro 366

 161 GCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCG 210
367 GluLysCysGlySerLysMetArgArgVal............LeuAlaGl 379

 211 CCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTACT 260
379 uLysCysAlaAspArgLysGlyArgPheSerAlaSerLysCysArgCysP 396

 261 TCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACG............AAA 298
396 heSerArg.........ProSerLysSerGlySerArgProGluLysArg 409

 299 TCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGA. 347
410 SerLysSerGlySerArgProGluLysArg.SerLysSerGlySerArgP 426

 348 ..............AAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATG 383
426 roGluLysArgSerGluSerGlyProArgProGlu......LysArgSer 440

 384 GACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCG 433
441 LysSerGlySerArgProGluLysCysGlySerLysMetArgArgValLe 457

 434 AGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCC 483
457 uAlaGluLysCysAlaAspArgLysGlyArgPheSerAlaSerLysCysA 474

 484 GACCCTAC.................GGTCATCATCAAAGAAGCCGCCGA 515
474 rgCysPheSerArgProSerLysSerGlySerArgProGluLysArgSer 490

 516 AGACTTCAAACGCGGCCT..............GTTGGTATTGAGCCGCC 550
491 LysSerGlySerArgProGluLysArgSerLysSerGlySerArgProGl 507

 551 TGACCGAACGTAAAAT.............CCATGTGTGTAAAGCAGCA 585
507 uLysArgSerLysSerGlySerArgProGluGluCysGly...SerAlaM 523
```

```
 586 GGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACACATGAATT 635
 523 etArgArgAlaLeuAlaGluLysCysAlaAspArgLysGlyArg..... 537

 636 TGGCGGCCCGCATCCTGCCGGCTT.....GAGTGGCACGCACA 673
 538 PheSerAlaSerLysCysArgCysPheSerArgProSerTrpSerGlyIl 554

 674 TTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAAT 723
 554 eLysProGluLysArgSerArgSer..... 562

 724 TATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAA 773
 563 .....GlySerArgProGluLys 568

 774 TACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAA...CAAACCGC 820
 569 ArgSerLysSerGly......SerArgProGluLysArgSerLysSerGl 583

 821 GCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGC 870
 583 ySerArgSerGluLysCysGly.....SerLysMetArgArgV 596

 871 GAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGG 920
 596 alLeuAlaGluLysCysAlaAspArgLysGlyArgPheSerAlaSerLys 612

 921 TGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGA 970
 613 CysArgCysPheSerArgProSerLysSerGlySerArgProGluLysAr 629

 971 TTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCG 1020
 629 gSerLysSerGlySerArgProGluLysArgSer..LysSerGlySerAr 645

1021 C...CGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCCATTT 1067
 645 gProGluLysArgSerLysSerGlySerArgProGlu..... 657

1068 CCTAAAAAACAAACTCTTCAAGTTCACGA......CAGCCGTCAACGGCG 1111
 658 .....LysArgSerLysSerGlySerArgProGluLysArgSerLysSer 672

1112 GCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTG 1161
 673 GlySerArgProGluLysArgSerLys..... 681

1162 GACATCCTGCCTACCTTGCTTTTGCGCGATTTAATCGTCGGCGATACCGA 1211
 682 .SerGlySerArgProGluIleCys.....GlySerAlaMet.... 693

1212 CAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACC..... 1255
 694 .....LysArgThrGluGly 698

1256 TCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTG 1305
 699 GluLysCysAlaArgLysAsnGlyArgPheAsnSerLys......ArgCy 713

1306 T...TGCGCAAAGTGCTGGAAACCA 1327
 713 sThrCysThrSerValGlyLysPro 721
```

seq_name: sp_rvirus:Q98WW2

seq_documentation_block:

```
ID   Q98WW2      PRELIMINARY;      PRT;  1504 AA.
AC   Q98WW2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GAG POLYPROTEIN [CONTAINS: CORE PROTEIN(S) P24].
GN   GAG.
OS   Avian leukosis virus.
OC   Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX   NCBI_TaxID=11864;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ENDOGENOUS AVIAN LEUKOSIS VIRUS 1, EV-1;
RX   MEDLINE=21165271; PubMed=11264350;
RA   Johnson J.A., Heneine W.;
RT   "Characterization of endogenous avian leukosis viruses in chicken
RT   embryonic fibroblast substrates used in production of measles and
RT   mumps vaccines.";
RL   J. Virol. 75:3605-3612(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ENDOGENOUS AVIAN LEUKOSIS VIRUS 1, EV-1;
RA   Johnson J.A., Heneine W.;
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO GAG GENE PROTEIN P24 (CORE NUCLEOCAPSID PROTEIN).
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
CC       KNOWN AS THE RETROPEPSIN FAMILY.
DR   EMBL; AY013303; AAK13200.1; -.
DR   HSSP; P03322; 2RSP.
DR   InterPro; IPR001969; Asp_protease.
DR   InterPro; IPR001995; Asp_prot_retrov.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR003308; Integrase_zn.
DR   InterPro; IPR004028; Retro_M.
DR   InterPro; IPR002156; RNaseH.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR000477; RVTse.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00607; gag_p24; 2.
DR   Pfam; PF02022; Integrase_Zn; 1.
DR   Pfam; PF02813; Retro_M; 1.
DR   Pfam; PF00075; rnaseH; 1.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF00077; rvp; 1.
DR   Pfam; PF00078; rvt; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
DR   PROSITE; PS50175; ASP_PROT_RETROV; 1.
KW   Aspartyl protease; Core protein; Hydrolase; Polyprotein;
KW   RNA-directed DNA polymerase; Zinc-finger.
FT   CHAIN         1    175       P19 MATRIX PROTEIN.
FT   CHAIN       178    239       P10.
FT   CHAIN       240    479       P27 CAPSID PROTEIN.
FT   CHAIN       489    577       P12 NUCLEOPROTEIN.
FT   CHAIN       578    701       P15 PROTEASE.
FT   CHAIN       709   1504       TRUNCATED REVERSE TRANSCRIPTASE-BETA
FT                                SUBUNIT.
FT   CHAIN       709   1280       REVERSE TRANSCRIPTASE-ALPHA SUBUNIT.
FT   CHAIN      1281   1504       TRUNCATED INTEGRASE.
SQ   SEQUENCE   1504 AA;  163199 MW;  979779BA89620435 CRC64;
```

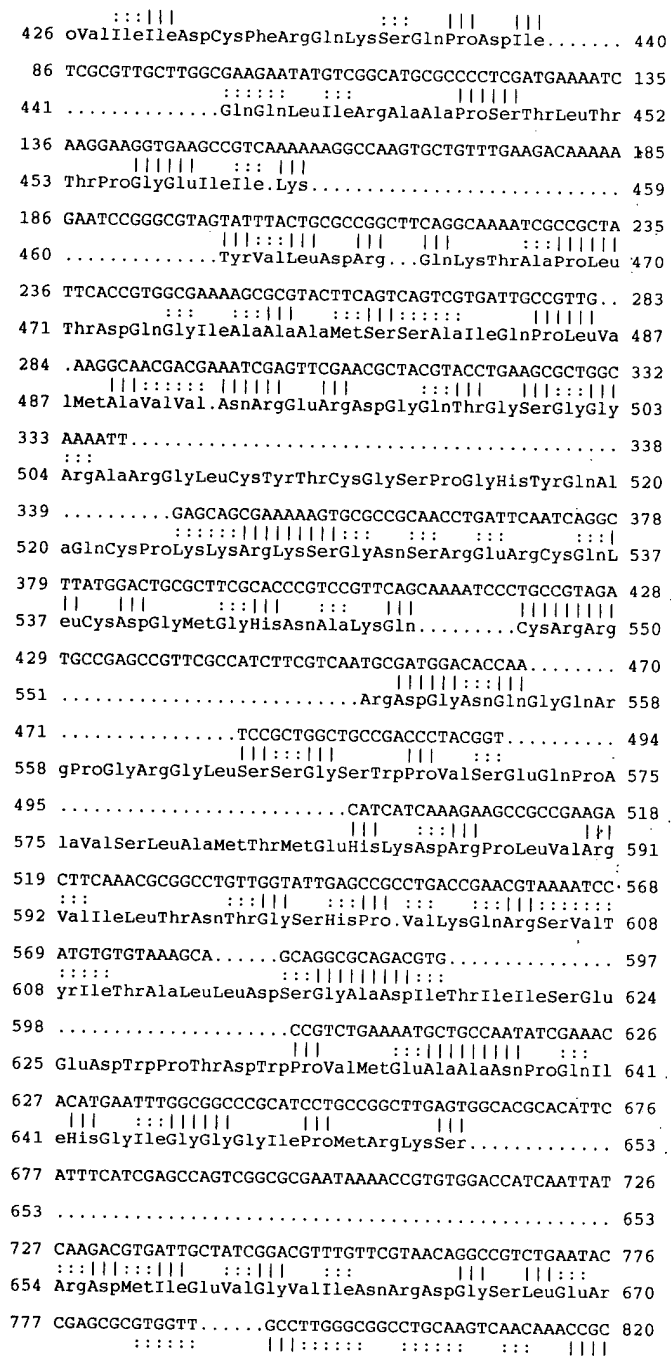
```
alignment_scores:
            Quality:  114.50              Length:    528
              Ratio:  0.504                 Gaps:     31
 Percent Similarity:  42.992   Percent Identity: 23.106

alignment_block:
 US-09-303-518D-131 x Q98WW2   ..

 Align seg 1/1  to: Q98WW2  from: 1  to: 1504

       1 ATGATTAAAATCAAAAAAGGTCTAAATCTGCCCATCGCGGGCAGACCGGA 50
     410 LeuIleLysAlaValGluGlySerAspLeuProProSerAlaArgAlaPr 426

      51 GCAAGTCATT.............TATGACGGCCCGGCCATTACCGAAG 85
```




```
seq_name: sp_invertebrate:Q9BKV7

seq_documentation_block:
ID   Q9BKV7      PRELIMINARY;      PRT;  1325 AA.
AC   Q9BKV7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PPG3.
GN   PPG3.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RA   Myler P.J., Sisk E., Cawthra J., Handley F., Vogt C., Robertson L.,
RA   McDonagh P., Ivens A., Nguyen D., Munden H., Stuart K., Worthey E.A.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC084329; AAK31375.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 6.
```

```
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00369; LRR_TYP; 5.
SQ   SEQUENCE   1325 AA;  133855 MW;  8898928BAEE9418F CRC64;


alignment_scores:
           Quality:  113.00            Length:    426
             Ratio:  0.546               Gaps:     12
 Percent Similarity:  48.592   Percent Identity: 20.657

alignment_block:
 US-09-303-518D-131/rev x Q9BKV7   ..

 Align seg 1/1  to: Q9BKV7  from: 1  to: 1325

   1337 TCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTCGTA 1288
    741 SerSerSerSerAlaProSerAlaSerSerSerSerAlaProSerSerSe 757

   1287 TTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATT 1238
    757 rSerSerAlaProSerAlaSerSerSerSerAlaProSerSerSerSerS 774

   1237 CCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCG 1188
    774 erSerAlaProSerAlaSerSerSerSerAlaProSerSerSerSerSer 790

   1187 CGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGT 1138
    791 AlaProSerAlaSerSerSerSerAlaProSerSerSerSerSerSerAl 807

   1137 GCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACT 1088
    807 aProSerAlaSerSerSerSerAlaProSerSerSerSerSerAlaPro. 823

   1087 TGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAG 1038
    823 ..................................................  823

   1037 TATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCC 988
    824 ......SerAlaSerSerSerSerAlaProSerSerSerSerSerAlaPr 838

    987 TTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCG 938
    838 oSerAlaSerSer..........................SerSerA 845

    937 CGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTTG 888
    845 laProSerSerSerSerSerAlaProSerAlaSerSerSerSerAlaPro 861

    887 TCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAA 838
    862 SerSerSerSerSerAlaProSerAlaSer...SerSerSerAlaProSe 877

    837 AACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAA 788
    877 rSerSerSerSerAlaProSerAlaSerSerSerSerAlaProSerSerS 894

    787 CCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCA 738
    894 erSerSerSerAla.........................ProSerAla 901

    737 ATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGG 688
    902 SerSerSer.........................SerAlaProSerSe 909

    687 CTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGC 638
    909 rSerSerSer......SerAlaProSerAlaSerSerSerSerAlaProS 924

    637 CAAATTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCG 588
    924 erSerSerSerSerSerAlaProSerAlaSerSer......SerSerAla 938

    587 CCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGCTCAATAC 538
    939 ProSerSer.............................SerSerSe 944

    537 CAACAGGCCGCGTTTGAAGTCTTCGGCGGCT...TCTTTGATGATGACCG 491
    944 rSerAlaProSerAlaSerSerSerSerAlaProSerSerSerSerSerA 961

    490 TAGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCG 441
    961 laProSerAlaSerSerSerSerAlaProSerSerSerSerSerSerAlaPro 977

    440 AACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAG 391
    978 SerAlaSerSerSerSerAlaProSerSerSerSerSerSerAlaProSe 994

    390 CGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGC 341
    994 rAla.....................................SerSerS 998

    340 TCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCG 291
    998 erSerAlaProSerSerSerSerSerAlaProSerAlaSerSerSerSer 1014

    290 TTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACG 241
   1015 AlaProSerSerSerSerSerAlaProSerAlaSerSerSerSerAlaPr 1031

    240 GTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCG 191
   1031 oSerSerSerSerSerAlaProSerAlaSerSer...SerSerAlaPro. 1046

    190 GATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCT 141
   1047 ..........SerSerSerSerSerAlaProSerAlaSerSerSerSer 1059

    140 TCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAA 91
   1060 AlaProSerSerSerSerSerAlaProSerAlaSerSerSerSerAl 1076

     90 CGCGACTTCGGTAATGGCCGGGCCGTCA 63
   1076 aProSerSerSerSerSerAlaProSer 1085

seq_name: sp_plant:O23054

seq_documentation_block:
ID   O23054      PRELIMINARY;      PRT;   402 AA.
AC   O23054;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   YUP8H12.26 PROTEIN.
GN   YUP8H12.26.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A., Osborne B.I., Vysotskaia V.S., Federspiel N.A.,
RA   Toriumi M., Yu G., Oji O., Araujo R., Chung E., Dewar K., Dietrich F.,
RA   Ecker J.R., Marziali A., Oefner P., Davis R.W.;
RT   "The sequence of YAC YUP8H12 from Arabidopsis thaliana chromosome 1.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A., Davis R.W., Federspiel N.A.;
```

```
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Theologis A.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC000098; AAB71465.1; -.
SQ   SEQUENCE   402 AA;  39721 MW;  D859630FB5C75466 CRC64;


alignment_scores:
            Quality:  112.50              Length:    438
              Ratio:  0.571                Gaps:     18
 Percent Similarity:  44.977   Percent Identity: 22.831

alignment_block:
 US-09-303-518D-131/rev x O23054   ..

 Align seg 1/1  to: O23054  from: 1  to: 402

   1331 TCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCG.........TATTC 1291
          |||         :::|||      |||||||:::|||           ||
     20 SerThrTyrProAsnThrTyrSerAsnSerAsnProAsnAlaAspAlaSe 36

   1290 GTATTTGCCCGGG...CAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
          |   :::|||:::   :::|||                ||||||||||
     36 rSerMetProSerThrAsnThrThrThrValProGlnThrSerSerSerS 53

   1243 CCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATT 1194
          ||:::|||:::         |||         |||   :::   |||
     53 erThrSerSerThrThrThrAlaThrGluSerSerSerGlyThrThrAla 69

   1193 AAATCGCGCAAAAGC...AAGGTAGGCAGGATGTCCAACGGCATTACGCG 1147
          :::|||   :::|||   |||   :::   ||||||:::::   |||::
     70 GluSerSerSerSerThrLysSerAlaThrMetSerGlySerThrThrHi 86

   1146 CTCATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTG 1097
          ::::                |||   :::   :::         ||||||
     86 sThr.............ThrSerSerAlaThrAlaSerSerThrAlaS 98

   1096 TCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGC 1047
          ::::::   :::|||
     98 erThrSerThrSerSer.............................. 103

   1046 GTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTT 997
               |||   ::::::   ::::::::::::   :::::::::
    104 .........TyrSerThrSerTyrSerThrSerSerThrLysThrThrTh 117

    996 GCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAAT 947
         :::      |||:::|||:::
    117 rMetThrGlySerThrIleSer........................... 124

    946 AATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATC 897
              :::   ::::::   ||||||   :::|||
    125 ........ThrThrAlaSerAlaAlaProThrSerThr............ 134

    896 ACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTT 847
              |||:::|||||||:::|||   ::::::|||   :::|||
    135 ......AlaSerThrSerThrSerSerTyrSerThrSerTyrSerThrSe 149

    846 CGCACCCAAAACGGTACGC.........AAGAGGCGCGGTTTGTTGACTT 806
         :::   ||||||:::             :::      |||     :::
    149 rSerThrLysThrThrThrValThrGlySerThrIleGlyThrThrAlaS 166

    805 GCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGA..... 768
              |||   :::|||:::   |||:::   :::
    166 erAlaAlaProThrSerThrSerThrSerThrAlaAsnSerSerAlaSer 182

    767 .........CGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTG 727
              :::|||   :::::::::::|||   |||:::|||:::
    183 SerThrThrAsnProSerSerGlySerLysProThrAlaMetThrGly.. 198

    726 ATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAAT 677
```

```
                :::::::|||          |||||||||:::|||
    199 ....ThrThrAlaAsnThrSerProSerAlaProThrSerSer..... 211

    676 GAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGT 627
                                                |||   ::::::
    212 ..................................ProSerThrThrAsn 216

    626 GTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTT 577
          |||      ||||||:::::::::::   |||   |||:::
    217 SerSerSerThrAlaAlaTyrThrSerSerGlySerLysProThr..... 231

    576 ACACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGC 527
                           :::|||   |||   :::::::|||
    232 ..................ThrValThrArgThrThrAlaAsn....... 239

    526 GTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCC 477
              ||||||:::||||||                  |||:::|||
    240 .....ThrSerSerSerAlaSerThr..............SerSerAla 249

    476 AGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATC 427
          |||                                 :::|||:::|||::||
    250 Ser.........................ProThrAsnSerSerThrSe 257

    426 TACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGC 377
          ||||                              ||||||      ||||
    257 rThrPro.................ThrAsnSerSerAlaGlySerLysP 268

    376 CTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGC 327
          ||                                            ::::::
    268 ro.....................................ThrThr 270

    326 GCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGGC 277
             :::|||||      :::::::|||   ::::::      |||:::||
    271 MetThrGlyThrThrThrAsnThrSerSerThrThrThrThrSerSerAl 287

    276 AATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCGA 227
          |   ||||||      |||:::
    287 aSerThrThrLysSerSerSer........................... 294

    226 TTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTTGTCT 177
                  ::::::|||:::|||::::::   |||         |||
    295 ..........SerSerAlaThrAsnSerSerSerGlySerLysProSer 307

    176 TCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATTTTCAT 127
          :::   ||||||              ||||||:::   |||
    308 ThrLeuSerThrThrThrAlaTyrThrAlaThrThrSerSerProThrAl 324

    126 CGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCGGTAA 77
          |||            |||   |||:::|||:::::::::::::::
    324 aGluProSerThrThrThrAlaSerLysProAlaThrSerSerThrProP 341

     76 TGGCCGGGCCGTCA 63
          |||   |||:::
    341 roAlaProProThr 345

seq_name: sp_bacteria:O06470

seq_documentation_block:
ID   O06470      PRELIMINARY;      PRT;   405 AA.
AC   O06470;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CEOA.
GN   CEOA.
OS   Burkholderia cepacia (Pseudomonas cepacia).
OC   Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC   Burkholderia.
OX   NCBI_TaxID=292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K61-3;
RA   Charron R.L., Barry J.J., Burns J.L.;
```

```
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97042; AAB58160.1; -.
DR   InterPro; IPR002215; HlyD.
DR   Pfam; PF00529; HlyD; 1.
SQ   SEQUENCE   405 AA;  44441 MW;  CCDC22ECFA35167F CRC64;


alignment_scores:
           Quality:  112.00            Length:    412
             Ratio:  0.647              Gaps:     23
 Percent Similarity:  41.990   Percent Identity: 26.456

alignment_block:
 US-09-303-518D-131 x O06470   ..

 Align seg 1/1  to: O06470  from: 1  to: 405

     103 GAATATGTCGGCATGCGCCCC.....................TCGATGAA 131
         |||   |||   :::|||||||                     ::::::::
      71 GluLysValAspValArgProGlnValSerGlyThrIleValAlaValAs 87

     132 AATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACA 181
         ::::|||:::|||      ||||||||||||:::|||||||||
      87 nPheLysAspGlyAlaLeuValLysLysGlyAspValLeuPheValIleA 104

     182 AAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCC 231
             :::|||      :::   :::      |||:::::::::::::|||
     104 spProArgProTyrGlnAlaGluValAspArgAlaAlaAlaGlnLeuAla 120

     232 GCTATTCAC.............................CGTGGCGAAAA 251
         |||   :::                             :::      ::
     121 AlaAlaGlnAlaArgAsnGlyTyrAlaGlnThrAspTrpGlnArgAlaGl 137

     252 GCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCG 301
         :|||::::::   :::   :::||||||   :::      |||||| :::
     137 nArgLeuIleGlyAspAsnAlaIleAlaLysArgAspTyrAspGlu.Lys 153

     302 AGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAA 351
                                     ||||||   ||| |||:::::
     154 ............................GlnAsnAlaAla.ArgGluA 160

     352 GTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCG 401
         ::      :::|||::::::
     160 laLysArgGluProGluGly.............................. 166

     402 TCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGC......CA 445
                               ||||||   ||||||   :::      ||
     167 .....................ArgArgSerArgAlaGlyAsnGlyAlaHi 176

     446 TCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCC.....*. 488
         |      :::      :::|||:::   ||||||      |||
     176 sGlnSerArgLeuTyrAlaHisHisArgAlaGlyValGlyProArgValA 193

     489 ..TACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTT 536
           |||:::|||   |||:::||||||             |||   |||
     193 laArgGlyAsnHisAlaArgGlnArgArg......ValGlyArgArgVal 207

     537 GGTATTGAGCCGCCTGAC.......................CGAACGTA 562
         |||          |||                         |||:::
     208 GlyArgAlaAlaAspAspAlaGlyIleGlyValAlaAspLeuArgVal.. 223

     563 AAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCT 612
                               ||||||||||||||||         :::
     224 ...................ValArgArgArgArgAlaAspLeuProAlaI 234

     613 GCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAG 662
           :::   |||::::::         ||||||   ::::::
     234 leHisGlnArgArgAla......GlnArgProGlnGlyAla......... 245

     663 TGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGT 712
                               ||||||      ||||||   ||||||::
     246 .................GlyArgAlaArgProArgGluArgAsnArgLe 256

     713 GGACCATCAATTATCAAGACGTGATTGCTAT...............CGGA 747
         :      |||      |||||||||   :::                 |||
     256 uLeuAlaGln......ArgArgAspArgPheGlyArgGlnProAlaArgH 271

     748 CGTTTGTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGG 797
               ||||||:::   |||          |||||||||
     271 isValValArgHisAspPro.........ArgAlaArgProLeuArgGln 284

     798 CGGCCTGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGA 847
         |||                     :::||||||      |||   |||
     285 ArgGlyArgHisProGlyProGlyProLeuArgThrArgGluGly..... 299

     848 AGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTG 897
                               |||:::   :::   ||||||      |||
     300 ................GlyArgGlnArgAlaAlaArgGlyAlaAlaArgA 311

     898 ATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTA 947
             |||         :::|||
     311 rgArgArgGlyAspGlnHisArg........................... 318

     948 TTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCA 997
                     |||   ::::::   |||   ||||||||||||
     319 ............ProGlyGlnGluValArgValArgArgArgPro..... 329

     998 AAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACG 1047
                                     |||:::   :::::::::
     330 .........................AlaGlyProArgValVal....... 335

    1048 CGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACGAC 1097
               ||||||         :::      :::   |||
     336 ........SerArgSerAlaAlaArgAspAlaAlaArgGlnProAlaArg 349

    1098 AGCCGTCAACGG.........CGGCGACCGCGCCATGGTACCGATCGGCA 1138
         :::|||:::|||         ||||||||||||   |||   :::|||||
     350 AspArgGluArgAlaValGlyArgArgProArg...GlyArgGluArgHi 365

    1139 CTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGC 1188
         |      |||   :::|||   |||:::   :::      |||   :::
     365 s......AlaAlaArgAlaSerGlyArgAlaGlyGluAlaAlaHisGlyP 380

    1189 GATTTAATCGTCGGCGATACCGACAGCGCGCA 1220
             :::   ||||||       ||||||
     380 roAspAspGlyArgArgCysAlaValArgAla 390

seq_name: sp_virus:Q98201

seq_documentation_block:
ID   Q98201      PRELIMINARY;      PRT;   579 AA.
AC   Q98201;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MC033L.
GN   MC033L.
OS   Molluscum contagiosum virus subtype 1 (MCVI).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Molluscipoxvirus.
OX   NCBI_TaxID=10280;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96325459; PubMed=8670425;
RA   Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA   Moss B.;
RT   "Genome sequence of a human tumorigenic poxvirus: Prediction of
RT   specific host response-evasion genes.";
RL   Science 273:813-816(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA   Moss B.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U60315; AAC55161.1; -.
DR   InterPro; IPR003600; Ig_like.
```

```
DR   SMART; SM00410; IG_like; 1.
SQ   SEQUENCE   579 AA;  62626 MW;  A37930DF92D311A3 CRC64;


alignment_scores:
            Quality:  112.00              Length:    473
              Ratio:  0.563                Gaps:     23
 Percent Similarity:  42.072   Percent Identity: 21.142

alignment_block:
 US-09-303-518D-131 x Q98201   ..

 Align seg 1/1  to: Q98201  from: 1  to: 579

     115 ATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAA..... 159
           |||||||||   :::   |||      :::   ::::::   :::
       1 MetArgProHisValLeuIleThrLeuAlaThrCysAlaLeuArgAlaLe 17

     160 .GGCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTG 208
          :::||||||:::   |||   :::   :::          |||   |||
      17 uAlaGlnValIleAspGluHisGluHisSerGluProProValSerThrT 34

     209 CGCCGGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCGTA 258
           |||      :::   :::::|||      |||:::::::::   :::
      34 rpProAspMetSerTyrLeuValAlaGluMetArgSerAspSerValLeu 50

     259 CTTCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAG..... 303
           ::::::::::          :::::|||:::::::::::::::
      51 MetArgGlyMet.........LeuAspGlyHisGluHisValArgCysTh 64

     304 .....TTCGAACGCTACGTACCTGAAGCGCTGG 331
                 :::|||    :::|||:::|||
      64 rCysValProLysCysAlaCysLeuGluProThrLeuProArgAlaAlaL 81

     332 CAAAATTGAGCAGCGAAAAAGTG..... 354
           :::   :::   :::::::|||
      81 euGluGlnAlaArgSerArgValLeuAspAlaHisAlaGlyArgValPro 97

     355 .....CGCCGCAACCTGATTCAATC 374
                          ||||||   ::::::   ::
      98 GlyLeuArgAlaProSerArgAlaAlaHisArgArgValValLeuThrAl 114

     375 AGGCTTA...TGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCTG 421
          :|||      :::|||            :::|||||||:::   :::   :
     114 aGlyCysArgPheThrGlnGlyPheProGluProPheGluGlyLeuTrpV 131

     422 CCGTAGATGCCGAGCCGTTCGCC..... 444
          ::::::::|||||||||   |||
     131 alAlaSerAlaGluProGlyAlaHisAspGluPhePheCysValGlyGlu 147

     445 .....ATCTTCGTCAATGCGATGGACACCAA 470
                          ||||||   :::|||   :::|||
     148 HisCysAspAlaHisLeuSerSerIlePheCysHisAlaAlaSerThrMe 164

     471 TCCGCTGGCTGCCGACCCT.....ACGGTCATCATCA 502
          |||||||||      |||                |||
     164 tProLeuAlaArgAlaProHisSerProProMetValThrPheSerAlaL 181

     503 AAGAAGCCGCCGAAGACTTCAAACGCGGCCTGTTGGTATTGAGCCGCCTG 552
          :::|||:::|||:::::   |||
     181 euSerAlaGlyGluAsnTyrLeuArg..... 189

     553 ACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCA..... 591
                       |||||||:::|||:::||||||
     190 .............LeuValCysArgAlaSerGlyAlaTyrProProVa 201

     592 .....GACGTGCCGT 601
                                          |||:::|||
     201 lAspThrLeuThrLeuValSerGlnProGlnGlnProGluAspAlaProC 218

     602 CTGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGC..... 642
          |||:::   |||:::   ::::::   :::   |||
     218 ysGluThrTyrAlaGlyThrAsnAlaAspSerThrGlyHisValGlyMet 234

     643 .....CCGCA 647
                                               ||
     235 AlaCysValArgSerAspAlaLeuAlaGlyAlaAlaCysAlaValGlnHi 251

     648 TCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATC..... 684
          |   :::      ::::::::::::|||   ::::::
     251 sArgGlyValThrThrSerAlaArgIleValLeuValProAlaAsnAspG 268

     685 ..GAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCAAGAC 732
                |||||||||         |||   :::      |||   |||
     268 lyAlaLysValGlyAlaTyrAlaAspValAspAlaAspPheTyrAlaAsp 284

     733 GTG.....ATTGCTATCGGACGTTT 752
          |||                :::|||:::      ||
     285 ValProProLeuProGluProGluSerAspSerLeuAlaValHisAlaLe 301

     753 GTTCGTAACAGGCCGTCTG.....AATACCGAGCGCGTGG 787
          |||||||:::|||:::                ::::|||  :::|||
     301 uPheValAlaGlyAsnThrGluLeuTyrValHisGlyThr.AlaAlaGly 317

     788 TTGCCTTGGGCGGCCTGCAAG...TCAACAAACCGCGCCTCTTGCGTACC 834
          :::|||   |||:::|||:::   :::|||:::|||::::::||||||
     318 ValProSerAlaSerCysArgCysAspThrArgArgCysThrCysVal.. 333

     835 GTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGG..... 877
                                      |||   ||||||:::|||
     334 .....LeuAlaProAlaThrTrpThrAlaG 342

     878 .....TTGACGCGGACAACCGCGTGATTTCCGGTTCGGTAT 913
                     |||:::|||::::::|||
     342 lyValValArgGluLeuAlaArgAlaAlaAla..... 352

     914 TGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCAC 963
                      |||         :::   :::   ||| :::
     353 ............HisAspLeuLeuLeuAlaValLeuAsp.ValHisA 364

     964 AATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTG 1013
           |||   :::      :::|||      :::|||      |||
     364 laSerGlyLeuAlaLeuAsnArgSerSerMetGln......ValTyrAla 378

    1014 GGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGGCC 1063
           |||:::   ||||||
     379 GluCysGlyProAlaGly..... 384

    1064 ATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCGGC 1113
                     :::   |||:::||||||:::::      ||||||
     385 .............ArgArgLeuArgValHisAsnThrGlyThrArgArg. 396

    1114 GACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTGGA 1163
                           :::::::::   |||||||:::|||
     397 ...................GlnArgValCysAlaArgGlyAlaCys... 405

    1164 CATCCTGCCTACCTT 1178
          :::||||||||||||
     406 GluProAlaTyrLeu 410

seq_name: sp_bacteriap:Q9RYP2

seq_documentation_block:
ID   Q9RYP2      PRELIMINARY;      PRT;   376 AA.
AC   Q9RYP2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ADENINE DEAMINASE-RELATED PROTEIN.
GN   DRA0268.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
```

```
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001863; AAF12376.1; -.
DR   TIGR; DRA0268; -.
KW   Complete proteome.
SQ   SEQUENCE   376 AA;  39845 MW;  7AB7FF32F8C45651 CRC64;


alignment_scores:
            Quality:  111.50            Length:    290
              Ratio:  0.961              Gaps:     15
 Percent Similarity:  40.000   Percent Identity: 26.897

alignment_block:
 US-09-303-518D-131 x Q9RYP2   ..

 Align seg 1/1  to: Q9RYP2  from: 1  to: 376

     639 CGGCCCGCA.............TCCTGCCGGCTTGAGTGGCACGCACA 673
         |||||||||               |||:::||||||      ||||||||
      35 ArgProAlaLeuArgGlyAspAspSerAlaArgLeuAlaGlyHisAlaHi 51

     674 TTCATTTCAT......CGAGCCAGTCGGCGCGAATAAAACCGTGTGGACC 717
         |     |||         |||:::|||
      51 sHisThrHisPheGlyAlaAlaGlyArgThrGly..............  62

     718 ATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCG 767
                                              :::||||||
      63 .....................................ThrGlyArgPro  66

     768 TCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAAC 817
         :::         |||   |||      |||           ::::::::
      67 AlaAlaGlyProAlaGlyGlyGlnSerGlyAlaGlyGluProArgArgVa 83

     818 CGCG...CCTCTTGCGTAC.........................CGTT 837
         :|||   |||      |||                          |||
      83 lAlaHisProAlaLeuTyrSerGlyGlyArgAlaAspArgAlaGlyArgA 100

     838 TTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGA 887
           |||      :::   :::   ||||||:::   :::|||      :::
     100 laGlySerAlaAlaGlyAlaArgTyrArgGluIleLeuGly...ProAla 115

     888 CAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTGCACAAGGCG 937
         :::||||||      |||      ::::::|||            |||
     116 ArgProArgGlyAspArgProTrpValSerArg......ArgLeuArgAl 130

     938 CGCATGATTATTTGGGACGCTA.......................... 959
          |||      ::::::||||||
     130 aAlaAlaArgLeuAlaThrLeuArgGlyAlaGlyAsnLeuArgGlyArgA 147

     960 ............................CCACAATCAGATTTCCGTTAT 980
                                     |||   :::      |||:::
     147 rgGlySerAlaAlaArgArgArgAspAlaProProAla......ArgTrp 161

     981 CGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGT.............. 1016
         ||||||||||||::         :::||||||||||
     162 ArgArgArgProArg.........AlaArgLeuGlyArgGlyHisLeuAr 175

    1016 ..................................................  1016

     175 gSerAlaArgAlaLeuAlaHisAlaProAspValSerArgProAspArgH 192

    1017 .....TGCGCCGCAGCCGGACAAATACTCCATCACGCGCACCACTCTCGG 1061
               :::::::::||||||||||      :::   |||   :::   |||
     192 isArgAlaCysGlyAlaGlyGlnArgArgArgProAlaGlyArgArgArg 208

    1062 CCATTTCCTAAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGCG 1111
         |||:::      ::::::::::::::::::   |||      ||||||
     209 ProLeuArgAlaArgArgValValGluLeuLeuAspValGlyGlnArgPr 225

    1112 GCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGCCGTTG 1161
                |||||||         :::|||      ||||||:::|||   |
     225 oAlaArgArgHisProGlyHisGlnHisSer...AlaArgArgAlaSerG 241

    1162 GACATCCTGCCTACCTT.........................GCTTTT 1184
         ||   ||||||                                :::
     241 lyGlyProAlaArgArgGlnArgArgGlyProAlaArgGlyGlySerGly 257

    1185 GCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCT 1234
         ||||||      ||||||||||:::|||
     258 AlaArgThrAlaArgArgArgHisArgPro..................... 267

    1235 TGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGC 1284
          |||   |||||||||   ||||||   |||         |||   |||:
     268 .GlyArgGlyArgArgGlyProArgAlaValAlaProAlaLeuArgGlyS 284

    1285 AAATACGAATACGGCCCGCT 1304
         ::         |||||||||
     284 er...AspAspArgProAla 289

seq_name: sp_invertebrate:Q9N5K0

seq_documentation_block:
ID   Q9N5K0      PRELIMINARY;      PRT;  1203 AA.
AC   Q9N5K0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 120.1 KDA PROTEIN.
GN   H43E16.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Miller N., Carter T.;
RT   "The sequence of C. elegans cosmid H43E16.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006669; AAF39909.1; -.
DR   InterPro; IPR000082; SEA.
DR   SMART; SM00200; SEA; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1203 AA;  120104 MW;  E04CADC2BA74713B CRC64;


alignment_scores:
            Quality:  111.50            Length:    512
              Ratio:  0.514              Gaps:     19
 Percent Similarity:  42.383   Percent Identity: 21.289

alignment_block:
 US-09-303-518D-131/rev x Q9N5K0   ..
```

```
Align seg 1/1  to: Q9N5K0  from: 1  to: 1203

1343 CAGCCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTA 1294
411 GlnGlySerProSerThrValSerGlnThr......ThrProGlyValSe 425

1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
425 rSerAlaSerThrGlyMetThrThrSerGlnAlaSerLeuArgSerThrG 442

1243 CCAATTCC......AAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCG 1200
442 lnAsnSerValThrSerThrProGlySerPheThrValThrSerThrPro 458

1199 ACGATTAAATCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTAC 1150
459 ThrIleThrSerThrGln.........GlyAlaMetAlaSerSerSerSe 472

1149 GCGCTCATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGG 1100
472 rAsnSer......ProThrSerThrGlnAla..................... 480

1099 CTGTCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTG 1050
480 ..................................................... 480

1049 CGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTC 1000
481 ..............IleSerSerSerThrValThrThrValAlaSerSe 492

999 TTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCA 950
492 rThrIleAlaProThrSerSerGlnGluSer..................... 502

949 AATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAA 900
503 .....SerSerIleAlaSerSerSerAlaProSerGlnSerThrThrVal 517

899 ATCACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACAC 850
518 IleThrValSerThrThrAlaThrValSerSerGlyGlnSerThrSerTh 534

849 C.............................TTCGCACCCAAAACGGTAC 830
534 rPheGlyThrThrIleGlyGlnGlySerSerPheGlySerSerThrIleG 551

829 GCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGC 780
551 lnThrThrGlnGlyThrSerSerPheAlaProIleProSerThrAlaGly 567

779 TCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTC 730
568 Ser.................SerSerGlnThrProGlySerMetSerSe 578

729 TTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGA 680
578 r.......................ThrGlyThrThrValGlyGlnMetS 587

679 AATGAATGTGCGTGCCACTCAAGCCGGCAGGATGCGGG............ 642
587 erSerSerPheGlnProThrAlaProThrSerLeuGlyThrIleMetThr 603

641 ...CCGCCAAATTCATGTGTT...TCGATATTGGCAGCATTTTCAGACGG 598
604 ThrProGlyThrSerSerIleProSerIleSerThrSerValAsnSerGl 620

597 CACGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGC 548
620 ySerSerThrIleGlySerThrValThrGlnAlaProSerSerSerThrS 637

547 GGCTCAAT................................ACCAAC 534
```

```
637 erMetGlyProSerProSerGlnSerThrAlaGlySerThrMetThrSer 653

533 AGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTC 484
654 AlaProProValThrSerSerSerSerAla.........AsnThrGlySe 667

483 GGCAGCCAGCGGA...TTGGTGTCCATCGCATTGACGAAGATGGCG.... 441
667 rThrSerSerGlyThrThrValSerValGlnThrThrGlnValSerSerT 684

441 ..................................................... 441

684 hrThrSerProValAlaSerSerSerSerGlnMetThrSerThrGlnGln 700

440 ...AACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGACGGGTGCG 394
701 ProSerGlySerSerSerSerIleGlySerThrValAsnGlnGlySerSe 717

393 AAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGC 344
717 rSerValThrThrGlnProPro.............................. 724

343 TGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCG 294
725 ..........AlaSerSerArgSerThrAlaSerGlnGlySerSerSer 737

293 TCGTTGCCTTCAACGGCAATCACGACTGACTGAAGTACG......CGCTT 250
738 AlaGlnProIleAlaSerSerSerThrMetGlySerThrAlaGlySerSe 754

249 TTCGCCACGGTGAATAGCG.................................. 231
754 rSerProGlnProThrAlaSerSerThrSerValProSerSerThrGlyA 771

231 ..................................................... 231

771 laThrSerSerGlySerThrValGlySerSerThrMetGlySerThrGln 787

230 GCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTT 181
788 SerSerLeuProSerSerThrMetThrAsnThrGly.............. 799

180 GTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATTT 131
800 .SerThrGlySerThrValThrAsnGlnLeuAlaSerSerSerThrTyrG 816

130 TCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCG 81
816 lyAlaSerSerThrGluProIleAlaSerSerThrAlaAsnProGlySer 832

 80 GTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGT 45
833 SerThrSerGlyGlnThrAlaValThrThrGlnGly 844
```

```
seq_name: sp_bacteria:P94909

seq_documentation_block:
ID   P94909      PRELIMINARY;      PRT;   529 AA.
AC   P94909;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 58.7 KDA PROTEIN (FRAGMENT).
OS   Microbacterium ammoniaphilum.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Micrococcineae; Microbacteriaceae; Microbacterium.
OX   NCBI_TaxID=42460;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15354;
RX   MEDLINE=96257250; PubMed=8654988;
RA   Striebel H.M., Seeber S., Jarsch M., Kessler C.;
RT   "Cloning and characterization of the MamI restriction-modification
```

```
RT   system from Microbacterium ammoniaphilum in Escherichia coli.";
RL   Gene 172:41-46(1996).
DR   EMBL; X79027; CAA55649.1; -.
KW   Hypothetical protein.
FT   NON_TER      1      1
SQ   SEQUENCE   529 AA;  58716 MW;  5DC1216237DF772D CRC64;


alignment_scores:
            Quality:  110.50            Length:    490
              Ratio:  0.531               Gaps:     25
 Percent Similarity:  42.449   Percent Identity: 24.694

alignment_block:
 US-09-303-518D-131 x P94909   ..

 Align seg 1/1  to: P94909  from: 1  to: 529

      18 AGGTCTAAATCTGCCCATCGCGGG..........................  41
         |||:::      |||:::||||||
      99 ArgAspAspArgAlaGluArgGlyArgArgGluProAlaArgProAlaAl 115

      42 .CAGACCGGAGCAAGTCATTTATGACGGCCCGGCCATTACCGAAGTCGCG  90
          :::   :::      ||||||       :::|||       |||
     115 aArgAspAlaArgArgHisLeuArgLeuGlyAlaHisProAlaArgArgG 132

      91 TTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAAGGA 140
           :::   ||||||      ||||||            |||   :::|||
     132 lyValLeuArgArgHisArgArgHisArgAlaArgGlyGlu...ArgGly 147

     141 AGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGAATC 190
         |||      ||||||:::   |||           ||||||
     148 ArgGlyProArgGlnGlnValPro............ArgGlnHisProAr 160

     191 CGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCTATTCAC 240
          |||                  |||:::::::|||        |
     160 gGly.....................ArgArgAspArgAlaGlyArgP 169

     241 CGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGCAA 290
         ||           |||           |||   :::|||   ||||||
     169 roGlyLeuHisArgArg.........ArgArgArgAlaArgArgArgGln 182

     291 CGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCTGGCAAAATTGA 340
            |||   :::||||||                  |||      ::
     183 GlyArgProGlnValArg.................HisGlyAspAspGl 193

     341 GCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCTTATGGACTGCG 390
         ::::|||   :::      :::|||   :::   |||         ::::
     193 nHisArgAlaAspProArgArgProArgAspProArg.........AlaG 207

     391 CTTCGCACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTT 440
         ::   ||||||     :::   :::      :::::::::|||  :::
     207 lyHisHisProLeuArgGluGlyGlyGlnGlyGlnGluAlaArgProLeu 223

     441 CGCCATCTTCGTCAATGCGATGGACACCAATCCGCTGGCTGCCGACCCTA 490
         |||           :::   |||   |||:::::         |||
     224 Arg............AlaArgGlyGlyGlnGlyGlyProGlyGlyProAl 236

     491 CGGTCATCAT.................CAAAGAAGCCGCCGAAGACTTC 522
          |||:::|||                 :::|||   |||   |||
     236 aGlyArgHisProAlaLeuGlyAspGluAspArgProArgAlaArgGlnG 253

     523 AAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCATGT 572
           :::|||
     253 lyAlaArg..................................... 255

     573 GTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCG 622
              :::   :::|||:::|||:::           |||   ||
     256 .........AlaAlaHisArgGluAlaLeuHisLeuArgLeuGlnArgAr 269

     623 AAACACATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCAC 672
         |:::           |||      :::||||||   :::   :::|||:
     269 gArg.........GlyArgAlaAspGlyCysArgAlaGlnGlyArgAlaA 283

     673 ATTCATTTCATCGAGCCAGT..............CGGCGCGAATAAAAC 707
         ::      :::||||||:::               |||||||||
     283 rgArgAlaArgArgAlaGlyGlyGlyGlyLeuProArgArgGlu...... 297

     708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCG 757
                        :::|||      ::::::|||          ||
     298 ..................AspArgValArgAlaHisArgProArgProAr 308

     758 TAACAGGCCGTCTGAATAC.........CGAGCGCGTGGTTGCCTTGGGC 798
         |:::|||   :::              ||||||   |||       :::|
     308 gGlyArgGlyGlyAlaAlaArgLeuAspArgAlaGlyGlyValGlyAlaA 325

     799 GGCCTGCA.......................................... 806
         ||||||||
     325 rgProAlaArgProArgArgLeuArgHisProArgProAlaAspLeuPro 341

     807 .........AGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGA 847
                  |||   :::|||   |||   :::      |||   ||
     342 HisGlyGlyProGlnGlyValAlaArgLeuAspHisProGlnGlyLeuGl 358

     848 AGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTG 897
         ||||::::::      ||||||                  :::|||:::
     358 uGlyAlaAlaGlySerArgArg..................HisProHisA 369

     898 ATTTCCGGTTCGGTATTGAA......CGGTGCGATTGCACAAGGCGCGCA 941
           :::|||      :::::      |||   |||      ||||||
     369 rgLeuArgAlaArgLeuHisGlnGlyArgGlyAspLeuLeuArgArg... 384

     942 TGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCC 991
                        |||:::   |||         ||||||   ||
     385 ..................ProArgArgAspArgLeuGlyArgArgGlyPr 395

     992 GCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCC 1041
         |:::|||                  |||   :::||||||   |||:
     395 oArgGlnGlyGln............GlyAlaHisGlyGlyGlnGlyLeuA 408

    1042 ATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTT 1091
         ::||||||   :::   ||| :::   :::   :::::::||| :::
     408 rgHisAlaGlyArgArgArg.ArgGlyValProSerGlnLeu.AlaValL 424

    1092 CACGACAG.........CCGTCAACGGCGGCGACCGCGCCATGGTACCGA 1132
           ||||||           ::::::|||:::   |||   |||
     424 euArgGlnGlyValCysGlnAlaSerAlaThrLeuAlaLeuTrpMetThr 440

    1133 TCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTT 1182
         |||:::   :::              |||       :::   |||   |||
     441 SerGlyArgLeu.........LeuArgGluLeuAlaGlyLeuLeuCysAr 454

    1183 TTGCGCGATTTAATCGTCGGCGATACCGACA................... 1213
             ::::::   |||             |||
     454 gAspSerValGluSerHisLeuAspValThrThrTrpArgAlaAspAlaS 471

    1214 .....GCGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCG 1258
              :::||||||   |||::::::   |||   |||:::      :::
     471 erIleSerArgArgThrTrpThrValMetAsn...ThrArgLeuArgGln 486

    1259 CTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTG 1308
                  :::   :::||||||      |||:::::::::
     487 AlaGlnLysSerLeuAlaAlaArgArgProThrThrSerSerSerIleSe 503

    1309 CGCAAAGTGCTGGAAA 1324
             :::||||||:::
     503 rArgSerCysTrpGln 508

seq_name: sp_human:Q9H9M1

seq_documentation_block:
ID   Q9H9M1      PRELIMINARY;      PRT;   639 AA.
AC   Q9H9M1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
```

```
DE   NT2RM4002189 PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK022723; BAB14204.1; -.
SQ   SEQUENCE   639 AA;  62709 MW;  5A198FA67808F0BE CRC64;


alignment_scores:
            Quality:  109.00              Length:    485
              Ratio:  0.507                Gaps:     21
 Percent Similarity:  44.330   Percent Identity: 23.093

alignment_block:
 US-09-303-518D-131/rev x Q9H9M1   ..

 Align seg 1/1  to: Q9H9M1  from: 1  to: 639

    1304 AGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAG 1255
           |||:::|||::::::   ||||||   :::|||:::
      95 SerAlaProPheThrSerLeuProPheSerThrSer..............  106

    1254 GTCTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCG.... 1209
           ||||||:::::::::::::::::::|||:::::::::::::|||||
     107 .SerSerAlaAlaSerThrSerAsnProAsnSerAlaSerLeuSerServ  123

    1208 .....GTATCGCCGACGATTAAATCGCGCAAA 1182
                        :::   ||||||   ::::::   :::
     123 alPheAlaGlyLeuProLeuProLeuProProThrSerGlnGlyLeuSer  139

    1181 AGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTG...CC 1135
           :::   :::      :::::::::|||   |||   |||   :::   ||
     140 AsnProThrProValIleAlaGlyGlySerThrProSerValAlaGlyPr  156

    1134 GATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCT..... 1098
           |:::|||      ::::::||||||   |||:::|||
     156 oLeuGly......ValAsnSerProLeuLeuSerAlaLeuLysGlyPheL  171

    1098 .....  1098

     171 euThrSerAsnAspThrAsnLeuIleAsnSerSerAlaLeuSerSerAla  187

    1097 ...........GTCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCC 1060
                   ::::::::::|||:::||||||   ::::::::
     188 ValThrSerGlyLeuAlaSerLeuSerSerLeuThrLeuGlnAsn.....  202

    1059 GAGAGTGGTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGC 1010
                              |||      :::|||:::   |
     203 ........................SerAspSerSerAlaSerAlaP  210

    1009 CGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGG 960
           |||||   :::         |||||||:::|||
     210 roAsn...LysCysTyrAlaProSerAlaIlePro..............  220

     959 TAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATAC 910
                 |||:::   :::   :::|||   ::::::|||   |||
     221 ...ThrProGlnArgThrSerThrPro...GlyLeuAlaLeuPheProGl  235

     909 CGAACCGGAAATCACGCGGTTGTCCGCGTCAACC..............  876
               |||:::   :::      |||:::||||||
     235 yProProSerProValAlaAsnSerThrSerThrProLeuThrLeuProV  252

     875 ..AATTCGCCG.........GCGGTAAGTTGAGACACCTTCGCACCCAAA 837
            :::||||||         |||:::|||   :::|||   ||||||
     252 alGlnSerProLeuAlaThrAlaAlaSerAlaSerThrSerAlaProVal  268

     836 ACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAAC 787
           :::      :::      :::||||||             |||   :::::
     269 SerCysGlySerSerAlaSerLeuLeuArgGlyProHisProGlyThrSe  285

     786 CACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCCGATAGCAA 737
           :        :::          |||:::::::::      |||:::
     285 rAspLeuHisIleSerSerThrProAlaAlaThrThrLeuProVal....  300

     736 TCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGC 687
                           |||:::   |||         :::||||||
     301 .................MetIleLysThrGluProThrSerProThrPro  311

     686 TCGATGAAATGAATGTGCGTGCCACTCAAG...CCGGCAGGATGCGGGCC 640
           |||                   :::|||   |||:::      |||
     312 Ser.................AlaPheLysGlyProSerHisSerGlyAs  322

     639 GCCAAATTCATGTGTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTG 590
           |||:::      :::      |||:::            ||||||   :
     322 nProSerHisGlyThrLeuGlyLeuSer............GlyThrLeuG  335

     589 CGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGGCTCAAT 540
           ::   |||      |||
     335 lyArgAlaTyrThrSerThr.............................  341

     539 ACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGT 490
                           |||         ||||||         ::
     342 .....................SerValProIleSerLeuSerAlaCysLe  351

     489 AGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGA 440
           ::::   |||   |||||||||   |||   :::              |
     351 uAsnProAlaLeuSerGlyLeuSerSerSerSerThrPro......LeuA  366

     439 ACGGCTCGGCATCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGC 390
           ||||||||         ::::::|||   |||       :::      |||
     366 snGlySerAsnProLeuSerSerIleSerLeuProProHisGlySerSer  382

     389 GCAGTCCATAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCT 340
           :::          |||           |||           ||||||
     383 ThrProIleAlaProValPheThrAlaLeuProSerPheThrSerLeuTh  399

     339 CAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGT 290
           |||            :::|||:::         |||         ||||
     399 rAsnAsnPheProLeuThrGlyAsnProSerLeuAsnProSerValSerL  416

     289 TGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCG...... 246
           |||||::::::   |||:::|||:::   ::::::      |||
     416 euProGlySerLeuIleAlaThrSerSerThrAlaAlaThrSerThrSer  432

     245 .........CCACGGTGAATAGCGGCGATTTTG.................. 222
                 |||         ||||||:::|||
     433 LeuProHisProSerSerThrAlaAlaValLeuSerGlyLeuSerAlaSe  449

     221 ....CCTGAAGCCGGCGCA.............................. 207
             |||   ::::::|||
     449 rAlaProValSerAlaAlaProPheProLeuAsnLeuSerThrAlaValP  466

     206 ........GTAAATACTACGCCCGGATTCTTTTTGTCTTCAAACAGCACT 165
                 ::::::|||   |||   :::   ||||||||   :::
     466 roSerLeuPheSerValThrGlnGlyProLeuSerSerSerAsnLeuSer  482

     164 TGGCCTTTTTTGACGGCTTCA......CCTTCCTTGATTTTCATCGAGGG 121
           :::|||   :::::::::|||      |||||||::
     483 TyrProGlyPheSerValSerAsnThrProSerValThrProAlaLeuPr  499

     120 GCGCATGCCGACATATTCTTCGCCAAGCAACGCGACTTCGGTAATGGCCG 71
           :::|||      ::::::||||||:::::::::::|||
     499 oSerPheProGlyLeuGlnAlaProSerThrValAlaAlaValThrProL  516
```

```
     70 GGCCG 66
        |||
    516 euPro 517

seq_name: sp_bacteria:Q93LB1

seq_documentation_block:
ID   Q93LB1      PRELIMINARY;      PRT;   299 AA.
AC   Q93LB1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE DNA-BINDING PROTEIN.
OS   Bifidobacterium animalis (Bifidobacterium lactis).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX   NCBI_TaxID=28025;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27536;
RA   Gonzalez Vara A., Rossi M., Altomare L., Gotti R., Matteuzzi D.;
RT   "Effect of pH and culture redox potential on the continuous culture of
RT   Bifidobacterium animalis ATCC 27536, studies on stability of
RT   recombinant plasmids and characterization of the insertion sequence
RT   IS1999.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ318089; CAC41624.1; -.
KW   DNA-binding.
SQ   SEQUENCE   299 AA;  34340 MW;  B71D96C9EC0A7A03 CRC64;


alignment_scores:
           Quality:  108.50            Length:    363
             Ratio:  0.743              Gaps:     21
 Percent Similarity:  40.220   Percent Identity: 24.793

alignment_block:
 US-09-303-518D-131 x Q93LB1   ..

 Align seg 1/1  to: Q93LB1  from: 1  to: 299

    189 TCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGC..... 230
        :::||||||   :::        |||   |||:::::|||
     18 AlaGlyArgProVal...LeuLeuGlyAlaArgAspArgArgGlyArgAl 33

    231 .CGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCC 279
         |||   :::|||   |||              |||   |||      |
     33 aArgValAlaProArgArg.............ValArgArgProValA 45

    280 GTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGTACCTGAAGCGCT 329
        ||   |||:::||||||                  |||      :::|||
     45 rgAspArgAspArgArgProAspProAlaLeuArgArg......GlnAla 59

    330 GGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTCAATCAGGCT 379
        ||||||         :::   :::::::|||         :::
     60 GlyLysThrAlaGlyHisGlyAlaGlyProPheArgProGlyAlaSerAr 76

    380 TATGGACTGCGCTTCGCACCCGTCCGTT.....................C 408
         ||||||:::::::::::::     |||                   :
     76 gMetAspAlaGlyAlaArgArgArgValHisLysThrArgArgGlyGlyA 93

    409 AGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGC 458
        ::::::::      ||||||   :::|||   |||::::::|||::::::
     93 rgHisGlyArgValArgArgLeuGluAlaGlyArgArgIleArgArgAla 109

    459 GATGGACACCAATCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAG 508
                              |||      ||||||:::
    110 Pro...................ArgGlyArgGlyHisArgProValPr 119

    509 CCGCCGAAGACTTCAAACGCGGCCT......GTTGGTATTGAGCCGCCTG 552
         ||||||         ||||||         |||      ||||||
    119 oArgArgAlaThrGlyArgArgProArgHisGlnGlyAlaLeuProPro. 135

    553 ACCGAACGTAAAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTC 602
                                  ::::::||||||:::|||   :::
    136 .......................AlaThrArgArgHisArgGlnAla 143

    603 TGAAAATGCTGCCAATATCGAAACACATGAATTTGGCGGCCCGCATCCTG 652
              ||||||   |||:::                     :::
    144 ......ArgCysGlnGlyArgGln...................AlaLe 151

    653 CCGGCTTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAAT 702
         ||||||:::
    151 uArgLeuGln..................................... 154

    703 AAAACCGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTT 752

    154 ..................................................... 154

    753 GTTCGTAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCC 802
                                                   :::|||||
    155 ..........................................AlaArgPr 157

    803 TGCAAGTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTG 852
        ||||:::|||:::      |||        |||      |||:::
    157 oAlaAspGlnArgArgIleProArgArgThrArg...ProCysGlnThrH 173

    853 TCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTC 902
          :::      |||:::              ||||||||||||
    173 isGlyProValArgHis............ProGlyGlnProArgProHis 185

    903 ...............................CGGTTCGGTATTGAACG 919
                                       |||:::   ::::::||
    186 AlaArgProArgArgLeuProLysGlyHisThrArgLeuProValArgAr 202

    920 GTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAG 969
        |   ::::::||||||||||:::                          |
    202 g...GlnAlaThrArgArgThr.......................HisA 210

    970 ATTTCCGTTATCGAAGAAGGCCGCAG..............CAAAGA... 1001
        ||   |||      ||||||||||:::            ||||||
    210 spAlaArgAlaAspArgArgProAspArgThrGlyAlaLeuGlnArgLeu 226

   1002 ...GCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGC 1048
           ::::::||||||            |||   :::         |||||
    227 ProGlyThrArgLeuProArgProArgAlaGlnLysThr......HisAl 241

   1049 GCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAA...GTTCACG 1095
        |:::|||   ||||||:::|||               |||   :::|||
    241 aArgHisProArgProLeuProProProProLeuLysGlnArgThrHisA 258

   1096 ACAGCCGTCAACGGCGG..............CGACCGCGCCATGGTACC 1130
          |||:::||||||:::              |||   |||||||||
    258 rgSerHisGlnArgGlnAlaArgAspProAlaArgGlnArgHisGlyLeu 274

   1131 GATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCT 1169
           :::|||        :::   ::::::   ||||||
    275 ArgGlnHisHisGlnLeuHisProThrMetProHisPro 287

seq_name: sp_bacteria:Q9KXM5

seq_documentation_block:
ID   Q9KXM5      PRELIMINARY;      PRT;   802 AA.
AC   Q9KXM5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE INTEGRAL MEMBRANE ATPASE.
GN   SCD12A.15C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
```

```
RA   Seeger K.J., Harris D.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL357524; CAB93406.1; -.
DR   InterPro; IPR001757; E1-E2_ATPase.
DR   InterPro; IPR001454; Hydrolase.
DR   InterPro; IPR000150; Hypothet_cof.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
SQ   SEQUENCE   802 AA;  85177 MW;  0C8BC48FE22D10E1 CRC64;


alignment_scores:
           Quality:  108.50              Length:    501
             Ratio:  0.474                Gaps:       27
 Percent Similarity:  45.709   Percent Identity: 23.154

alignment_block:
 US-09-303-518D-131 x Q9KXM5   ..

 Align seg 1/1  to: Q9KXM5  from: 1  to: 802

      64 GACGGCCCGGCCATTACCGAAGTCGCGTTGCTTGGCGAA...GAATATGT 110
         ||||||       |||   |||      |||   ||||||   :::  ||
     159 AspGlyLeuGluIleAspGluSerLeuLeuThrGlyGluAlaAspProVa 175

     111 CGGCATGCGCCCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAG 160
         |      :::|||:::   ::::::   :::|||:::   |||      |
     175 lValLysGlnProGlyAspGlnValMetSerGlySerPheValValAlaG 192

     161 GCCAAGTGCTGTTTGAAGACAAAAAGAATCCGGGCGTAGTATTTACTGCG 210
         ||         |||  |||      :::   ||||||:::   |||:::|
     192 lyGlyGlyAlaPhe..ArgAlaThrLysValGlyArgGluAlaTyrAlaA 208

     211 CC.................................................. 212
         ||
     208 laGlnLeuAlaGluGluAlaSerArgPheThrLeuValGlnSerGluLeu 224

     213 ......GGCTTCAGGCAAAATCGCCGCTATTCACCGTGGCGAAAAGCGCG 256
               |||:::           :::|||      |||      :::
     225 ArgSerGlyIleSerThrIleLeuLysTyrValThrTrpMetMetValPr 241

     257 TACTTCAGTC...........AGTCGTGATTGCCGTTGAAGGCAACGAC 294
          |||::::::           |||::: :::   ::::::   ||||||
     241 oThrAlaIleGlyLeuIleLeuSerGln.LeuLeuAlaLysAspAsnAsp 257

     295 ...GAAATCGAGTTCGAACGC.....................TACGTACC 320
            :::   ::::::   |||                      |||||
     258 LeuAspAspSerIleAlaArgThrValGlyGlyIleValProMetValPr 274

     321 TGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACCTGATTC 370
         ||||:::|||:::   |||:::|||              ::::::::::
     274 oGluGlyLeuValLeuLeuThrSerValAlaPheAlaIleGlyValValA 291

     371 AATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAAATCCCT 420
         ::   |||        :::|||                :::::::::|||
     291 rgLeuGlyArgLysGlnCysLeu...........ValGlnGluLeuPro 303

     421 GCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAA 470
         |||::::::::::                  :::      :::||||||
     304 AlaIleGluGly.................LeuAlaArgValAspThr.. 313

     471 TCCGCTGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACT 520
              ::::::   |||   |||      :::   |||::::::   |||
     314 ....ValCysLeuAspLysThrGlyThrLeuThrGluGlyGlyMetAspV 329

     521 TCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCGAACGTAAAATCCAT 570
                 |||   ||||||:::          ||||||   ::::::
     329 al...ThrAspLeuArgValLeuAspGlyAlaAspGluArgHisValArg 344

     571 ...GTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCT..... 612
            |||      |||   ||||||:::   |||      ||||||
     345 AlaValLeuGlyAlaLeuGlyAlaSerAspProArgProAsnAlaSerLe 361

     613 ....GCCAATATCGAAACACATGAATTTGGCGGC................ 642
            |||   |||:::|||:::          |||
     361 uGlnAlaIleIleAspThrTyrProAspAspGlySerAlaGlyAlaTrpA 378

     643 .................CCGCATCCTGCCGGCTTGAGTGGCACGCACATT 675
                         |||      ::::::   :::   :::
     378 rgAsnThrArgSerLeuProPheSerSerAlaArgLysTyrSerGlyAla 394

     676 CATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTA 725
         |||   |||      ||||||     |||  |||   ::::::
     395 ThrPheAspGluGlyAspGlyAlaAlaGlyThr...TrpLeuLeuGlyAl 410

     726 TCAAGACGTGATTGCT.........ATCGGACGTTTGTTCGTAACAGGCC 766
         |||||||:::                        |||      |||||||
     410 aProAspValLeuLeuProGluAspAspProAlaLeuAlaGluThrGlyA 427

     767 GTCTGAATACCGAG......CGCGTGGTTGCCTTGGGC............ 798
         ||||||||   :::      |||||||:::   |||:::
     427 rgLeuAsnGluGlnGlyLeuArgValLeuLeuLeuAlaArgAlaAlaArg 443

     799 .....................................GGCCTGCA 806
                                             :::|||
     444 AspLeuAspAspProGluValAlaGluGlyValGluProThrAlaLeuVa 460

     807 AGTCAACAAACCGCGCCTC.........TTGCGTACCGTTTTGGGTGCGA 847
          |||   :::   ||||||                  :::|||
     460 lValLeuGluGlnArgLeuArgProAspAlaAlaAspThrLeuArgTyrP 477

     848 AGGTGTCTCAACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTG 897
          ::::::|||   :::   :::::::::::::::::::|||||   |||
     477 heAlaAspGlnAspValArgAlaLysValIleSerGlyAspAsnAlaVal 493

     898 ATTTCCGGTTCGGTA..............TTGAACGGTGCGATTGCACA 932
                 |||:::|||            |||:::||||||   :::::
     494 SerValGlyAlaValAlaAlaLysLeuGlyLeuSerGlyAlaThrValAs 510

     933 AGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCG 982
         :   |||:::         ::::::   :::::::::   ::::::::::
     510 p...AlaArgArgLeuProAlaGluArgGluGluMetAlaGlyAlaLeuA 526

     983 AAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGAC 1032
         ::||||||            :::||||||   |||:::||||||
     526 spGluGly.........ThrValPheGlyArgValThrProGlnGlnLys 539

    1033 AAATACTCCATCACGCGC...ACCACTCTCGGCCATTTCCTAAAAAACAA 1079
         :::      ::::::         :::   ||||||
     540 ArgAspMetValAlaAlaLeuGlnSerArgGlyHis.............. 551

    1080 ACTCTTCAAGTTCACGACAGCCGTCAACGGCGGCGACCGCGCCATGGTAC 1129
                    |||:::|||:::::|||
     552 .............ThrValAlaMetThrGlyAsp................ 558

    1130 CGATCGGCACTTATGAGCGCGTAATGCCGTTGGACATCCTGCCTACCTTG 1179
                                           :::      :::|||
     559 ...............................GlyValAsnAspValLeu 564

    1180 CTTTTGCGCGATTTAATCGTCGGC..............GATACCGACAG 1214
```

```
          |||:::|||      :::|||                        ::::::::
     565 AlaLeuLysAspAlaAspIleGlyValAlaMetGlySerGlySerGluAl 581

    1215 CGCGCAGGCTTTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCT...T 1261
         :::::::|||::::::   :::   |||   :::::::::::::|||   |
     581 aThrArgAlaValAlaGlnIleValLeuLeuAsnAsnSerPheAlaThrL 598

    1262 TGTGCAGCTTCGTCTGCCCGGGCAAATACGAATACGGCCCGCTGTTGCGC 1311
         ||   |||   |||:::   |||:::          |||   :::   |||
     598 euProSerValValAlaGluGlyArgArgValIleGlyAsnIleThrArg 614

seq_name: sp_human:Q9Y2N2

seq_documentation_block:
ID   Q9Y2N2      PRELIMINARY;      PRT;  1294 AA.
AC   Q9Y2N2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   WUGSC:H_DJ0953A04.3A PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99063792; PubMed=9847074;
RA   Sulston J.E., Waterston R.;
RT   "Toward a complete human genome sequence.";
RL   Genome Res. 8:1097-1108(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cordes M., Bauer C., Holmes A.;
RT   "The sequence of Homo sapiens PAC clone RP5-953A4.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006014; AAD28063.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   1294 AA;  131942 MW;  6FC9761CC44162C2 CRC64;


alignment_scores:
            Quality:  108.50              Length:    524
              Ratio:  0.478                 Gaps:     22
 Percent Similarity:  43.321   Percent Identity: 21.565

alignment_block:
 US-09-303-518D-131/rev x Q9Y2N2   ..

 Align seg 1/1  to: Q9Y2N2  from: 1  to: 1294

    1310 CGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAA 1261
         :::|||||||:::      |||      ||||||::::::      :::||
     467 GlnAsnSerAsnSerGlnSer...ThrProGlySerSerGlyGlnArgLy 482

    1260 AGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGT 1211
         |   :::   :::        |||::::::   :::      :::|||
     482 sArgLysValGlnLeuLeuProSerArgArgGlyGluGlnLeuThrLeuP 499

    1210 CGGTATCGCCGACGATT.............................. 1194
              |||   :::
     499 roProProProGlnLeuGlyTyrSerIleThrAlaGluAspLeuAspLeu 515

    1193 ..........................AAATCGCGCAAAAGCAAGGT 1174
                                    ::::::   ::::::|||
     516 GluLysLysAlaSerLeuGlnTrpPheAsnGlnAlaLeuGluAspLysSe 532

    1173 AGGCAGGATGTCCAACGGCATTACGCGCTCA................... 1143
              ||||||::::::|||::::::
```

```
     532 rAspAlaAlaSerAsnSerValThrGluThrProProThrThrGlnProS 549

    1142 ..........TAAGTGCCGATCGGTACCATGGCGCCGGTCGCCGCCGTTG 1104
                    :::|||   ::::::   |||   |||||||||
     549 erPheThrPheThrLeuProAlaAlaAlaThrAla...SerProProThr 564

    1103 ACGGCTGTCGTGAAC..............TTGAAGAGTTTGTTTTTTAG 1069
         :::   ::::::                 |||:::|||||||
     565 SerLeuLeuAlaProSerThrAsnProLeuLeuGluSerLeuLysLysMe 581

    1068 GAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCG 1019
          :::   |||      :::                    :::|||:::::::
     581 tGlnThrProProSerLeuProProCysProGluSerAlaGlyAlaAlaT 598

    1018 CAACCCAGCCGAACAGC..............TCTTTGCTGCGGCCTTCT 984
         ::|||:::      |||              |||||||||   |||
     598 hrThrGluAlaLeuSerProProLysThrProSerLeuLeuProProLeu 614

     983 TCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCC 934
         ::::::::::::                                  :::||
     615 GlyLeuSerGln..............................SerGlyPr 621

     933 TTGTGCAATCGCACCG.........TTCAATACCGAACCGGAAATCACGC 893
         |   ::::::   |||         |||::::::::::|||       |||
     621 oProGlyLeuLeuProSerProSerPheAspSerLysProProThrThrL 638

     892 GGTTGTCC.........GCGTCAACCAATTCGCCGGCGGTAAGTTGAGAC 852
           |||:::          |||   ::      ||||||:::      |||
     638 euLeuGlyLeuIleProAlaProSerMetValProAlaThr......Asp 652

     851 ACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAG 802
         |||   ||||||   |||::::::   :::             |||:::
     653 ThrLysAlaProProThrLeuGlnAlaGlu............ThrAlaTh 665

     801 GCCGCCCAAGGCAACCACG.................CGCTCGGTATTCA 770
             |||:::||||||:::                 ::::::   |||
     665 rLysProGlnAlaThrSerAlaProSerProAlaProLysGlnSerPheL 682

     769 GACGGCCTGTTACGAACAAACGTCCGATA..............GCAATC 735
               :::   |||       |||                    |||
     682 euPheGlyThrGlnAsnThrSerProSerSerProAlaAlaProAlaAla 698

     734 ACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTC 685
         :::|||         |||         ::::::   ||||||      ||
     699 SerSerAlaSerProMetPheLysProIlePheThrAlaProProLysSe 715

     684 GATGAAATGAATGTGCGTGCCACTCAAGCCG.................. 654
         |   |||         :::|||      |||
     715 rGluLysGluGlyProThrProProGlyProSerValThrAlaThrAlaP 732

     654 ..................................................... 654

     732 roSerSerSerSerLeuProThrThrThrSerThrThrAlaProThrPhe 748

     653 ............GCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATT 616
                     ::::::   |||||||||   |||   |||   :::
     749 GlnProValPheSerSerMetGlyProProAlaSer...ValProLeuPr 764

     615 GGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGA 566
          |||   |||          |||::::::||||||:::      |||
     764 oAlaProPhePheLysGlnThrThrProAlaThrAlaProThr.... 779

     565 TTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAG... 519
                                 |||:::   |||   :::
     780 ........................ThrThrAlaProLeuPheThrGly 787

     518 ...TCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAGCGG 472
            :::||||||:::|||   :::      :::   ||||||:::   ::
     788 LeuAlaSerAlaThrSerAlaValAlaProIleThrSerAlaSerProSe 804

     471 ATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGC......TCGGCAT 428
         :      |||   :::          ::::::   |||      |||:::|
     804 rThrAspSerAlaSerLysProAlaPheGlyPheGlyIleAsnSerValS 821
```

```
     427 CTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAG 378
         ||:::::::::::::      :::      :::   :::|||:::   :::
     821 erSerSerSerValSerThrThrThrSerThrAlaThrAlaAlaSerGln 837

     377 CCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAG 328
         |||                      |||:::      :::   :::::
     838 Pro.................PheLeuPheGlyAlaProGlnAlaSerAl 848

     327 CGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTG......CCTT 284
         :|||||   |||        :::||||||   ::::::     |||
     848 aAlaSerPheThrProAlaMetGlySerIlePheGlnPheGlyLysProP 865

     283 CAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATA 234
           :::      ||||||         |||   ||||||
     865 roAlaLeuProThrThrThrThrValThrThrPheSerGlnSer..... 879

     233 GCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACG.........%. 195
               ||||||   |||      :::   |||:::
     880 .........LeuProThrAlaValProThrAlaThrSerSerSerAlaAl 893

     194 .......CCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGA 152
               ||||||      :::      ::::::   |||      :
     893 aAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAlaThrS 910

     151 CGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCT 102
         ::::::::|||:::|||   |||   :::         ||||||::::::
     910 erSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPheAsn 926

     101 TCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGTCATAAATGACTTG 52
            |||   ::::::::::::            ||||||      ::
     927 IleProPheGlySerSerAlaLysSerProLeuProSerTyrProGlyAl 943

      51 CTCCGGTCTGCCCGCGATGGGC 30
         ::::      ||||||:::|||
     943 aAsnProGlnProAlaPheGly 950

seq_name: sp_invertebrate:P91365

seq_documentation_block:
ID   P91365      PRELIMINARY;      PRT;  2232 AA.
AC   P91365;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   K06A9.1 PROTEIN.
GN   K06A9.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Geisel C., Gattung S.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC       OF THE A FORM.
DR   EMBL; U80846; AAC70889.1; -.
DR   EMBL; U80846; AAC70890.1; -.
KW   Alternative splicing.
FT   VARSPLIC    842    866       VASSSPAPSTSQNPNPSTSSGSSMI -> LATTSAPKPSVT
FT                                CLFMYD (IN ISOFORM B).
FT   VARSPLIC    870   1051       PYPSQSTSPVESSTTPSPGSPGTTLTSTSPSPSQSTTIGST
FT                                QGSTSPGISTTSEEMTSQGSTQTPGSTGSTVTQPSTVSDST
FT                                SSGSTVTVGSTEGSSSPIPSTSQNTNPSTSSGSSMSTQTPQ
FT                                SSQSTSPVESSTSGATSSSGSPGTTLTSISPSPSPSSTIGS
FT                                SQGSTSPVVSTISQGSTE -> KEIDQTAINTYKTYFNFAL
FT                                LVASKLNNESILTGYIDNFGYSAGLNDHQYYPTDDYNGIKS
FT                                VPFPIDGTDDDIDLDLKDVDKSLATADWTPPVADQTCMIFI
FT                                SAAPEDEYGGTTIKSTYTYFETVVGVLVGGAKSIPGLSIDK
FT                                NIVITNNTMNDRDASAVVSKLLELLPTA (IN ISOFORM
FT                                B).
SQ   SEQUENCE   2232 AA;  213840 MW;  08D69FA638E14CC8 CRC64;


alignment_scores:
            Quality:  108.50              Length:    453
              Ratio:  0.522                Gaps:     16
 Percent Similarity:  45.916   Percent Identity: 20.530

alignment_block:
 US-09-303-518D-131/rev x P91365   ..

 Align seg 1/1  to: P91365  from: 1  to: 2232

    1340 CCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCG..... 1296
         ||||||         :::   |||||||::      |||||||||
     502 ProSerSerProGlyThrMetSerThrVal......SerGlyProThrGl 516

    1295 .......TATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGT 1253
                 :::   :::||||||::::::                    |
     516 ySerThrValThrValValProGlySerSerThrSerProAlaProSerS 533

    1252 CTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCG 1203
         |||||   :::   |||:::::|||   :::   :::   ::::::
     533 erSerProAsnProSerSerSerProAlaSerThrGlySerThrIle... 548

    1202 CCGACGATTAAATCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCAT 1153
            ||||||::::::   :::|||   :::             :::|||
     549 ...ThrIleSerGlySerSerSerIleIleValSerThrValSerGlySe 564

    1152 TACGCGCTCATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGA 1103
          |||   |||              |||:::   :::   |||
     564 rThrValSer............GlySerThrGlyThrSerGlnSerThrL 577

    1102 CGGCTGTCGTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTG 1053
            |||     :::
     577 euAlaSerSerThrAla.................................. 582

    1052 GTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAG 1003
                                  |||   ::::::|||   |||:::||
     583 .................ThrProGlySerSerSerThrValProSerSe 593

    1002 CTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTC 953
         ||||      :::|||||||||   :::
     593 rSerSerProGlnProSerSerGlnSerPro..................... 603

     952 CCAAATAATCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAACCG 903
                                        ||||||   :::   :::
     604 ............................AlaProAsnThrGlySerThr 610

     902 GAAATCACGCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGA 853
               ::::::   |||:::::::::   ||||||:::::::::   ::
     611 ThrProSerGlnThrSerSerGlnSerProSerProSerMetAsnProSe 627

     852 CACCTTCGCACCC..............AAAACGGTACGCAAGAGGCGCG 818
         ::::   :::|||               :::|||:::      :::   :
     627 rSerSerThrProThrGlySerSerGlnSerThrIleThrProGluGlyS 644

     817 GTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTC... 771
         ::      |||:::      |||   ::::::|||   |||:::|||
     644 er......ThrAlaSerSerProThrGlySerThrGlySerThrPheSer 658

     770 ...AGACGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATA 724
                    ||||||::::::      :::
     659 ValAlaThrGluValThrSerGlnSerThrVal.................. 669

     723 ATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAA 674
                                        |||:::||||||
     670 ............................ProSerGlySer.......... 673

     673 TGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCG...CCAAATTCATGT 627
               :::      :::   ::::::   :::|||   |||:::|||
     674 ..SerLeuGlyThrGlnSerThrAsnSerSerProSerProSerSerLeu 689
```

```
    626 GTTTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTT 577
                  :::::::::|||      ||||||   |||:::
    690 SerProSerThrSerGlyMetSerThrLeuThrSerGluProSerProSe 706

    576 ACACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGC 527
        |||           :::         |||:::|||      |||:
    706 rSerThrGlnSerSerGlyAlaGlnSerThrLeuThrThrProSerProA 723

    526 GTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCC 477
        ::   ::::::|||:::::|||||      :::::::::::|||:::
    723 snProSerGlnSerThrSerSerLeuGluSerSerThrSerGlyAlaThr 739

    476 AGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATC 427
        ::::::      |||   :::   |||      :::   :::::::::||
    740 ThrSerSerGlySerAlaGlyThrThrMetThrSerProSerGlnSerSe 756

    426 TACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGC 377
        |:::::|||        ::::::       :::|||
    756 rSerValGlySerSerGlnGlySerThrSerProAlaAla.......... 769

    376 CTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGC 327
                                                  ::::::
    770 ..............................................SerThr 771

    326 GCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGGC 277
        :::||||||        |||:::::::   ::::::   |||:::::::::
    772 ThrSerGlyGluMetThrSerGlnGlySerThrGlnThrProGlySerSe 788

    276 AATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCGA 227
        :::::::|||:::                              |||||||
    788 rValSerThrSer..............................AlaAlaI 795

    226 TTTTGCCTGAAGCC.....................GGCGCAGTA 204
        |||||   ::::::                     |||:::::
    795 leLeuThrSerThrGlnGlnSerValSerThrAsnSerProGlySerThr 811

    203 AATACTACGCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTT 154
           |||   |||:::         :::|||:::|||:::          ::
    812 ValThrArgProSerThrValSerGlySerThrSerSerGlySerThrVa 828

    153 GACGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATT 104
        :|||::::::   :::          :::|||         :::   |
    828 lThrValGlySerThrGluAlaSerThrSerGlySerSerValAlaSerS 845

    103 CTTCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGTCATAAATGACT 54
        |||||||||:::   :::||||||         :::|||||      |||
    845 erSerProAlaProSerThrSerGlnAsnProAsnProSer......Thr 859

     53 TGCTCCGGT 45
        ||||||
    860 SerSerGly 862
```

```
seq_name: sp_bacteria:Q9RFF2

seq_documentation_block:
ID   Q9RFF2      PRELIMINARY;      PRT;   569 AA.
AC   Q9RFF2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   UREC.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=2.4.1;
RA   Choudhary M., Kaplan S.;
RT   "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT   sphaeroides 2.4.1.";
RL   Nucleic Acids Res. 0:0-0(2000).
DR   EMBL; AF195122; AAF24255.1; -.
DR   HSSP; P18314; 1FWF.
DR   InterPro; IPR001924; Urease.
DR   Pfam; PF00449; urease; 2.
DR   Pfam; PF02802; urease_C; 2.
DR   ProDom; PD002467; Urease; 2.
DR   PROSITE; PS01120; UREASE_1; 1.
SQ   SEQUENCE   569 AA;  61425 MW;  F543A679D3BBF1DA CRC64;


alignment_scores:
            Quality:  108.00              Length:    373
              Ratio:  0.643                 Gaps:     20
 Percent Similarity:  45.040   Percent Identity: 22.788

alignment_block:
 US-09-303-518D-131 x Q9RFF2   ..

 Align seg 1/1  to: Q9RFF2  from: 1  to: 569

    121 CCCTCGATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCT 170
        |||:::   :::|||      ||||||            |||:::::||
    116 ProGlyThrGluIleIleAlaGlyGlu...........GlyArgIleLe 128

    171 G.............................................T 172
        |                                            :
    128 uThrAlaGlyGlyMetAspAlaHisIleHisPheIleCysProGlnGlnI 145

    173 TTGAAGACAAAAAGAATCCGGGCGTAGTATTT.............ACT 207
        ::||||||:::   :::   |||::::::                |||
    145 leGluAspSerLeuHisSerGlyIleThrThrMetLeuGlyGlyGlyThr 161

    208 GCGCCGGCTTCAGGCAAAATCGCCGCT.................AT 236
        :::||||||   |||   :::|||:::
    162 GlyProAlaHisGlyThrLeuAlaThrThrLeuHisAlaGlyAlaLeuAl 178

    237 TCACCGT...........GGCGAAAAGCGCGTACTTCAGTCAGTCGTGA 274
         ||||||           |||   :::|||:::   ::::::      :
    178 aHisArgAlaAspAlaAlaGlyGlyArgArgLeuProAspGlnProArgL 195

    275 TTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAA.............. 309
        ::      :::|||:::   :::         :::
    195 euArgGlyGlnGlyGlnCysGlnProAlaArgArgAlaArgGlyAlaGly 211

    310 .........CGCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAA 350
                |||   ||||||||||||        |||:::   :::
    212 ProArgGlyArgLeuValProGluAlaAlaArgGlyLeuGlyHisHisAl 228

    351 AGTGCGCCGCAACCTGATTCAATCAGGCTT.......ATGGACTGCGCTT 393
              ::::::|||:::      |||               ||||||:
    228 aCysGlyHisArgLeuLeuProValGlyArgArgProHisGlyCysAlaG 245

    394 CG.........CACCCGTCCGTTCAGCAAAATCCCTGCCGTAGATGCCGA 434
        ::          |||      ::::::::::::      ||||||   :::
    245 lyAspAspProHisArgHisAlaGluArgGluArgLeuArgArgGluHis 261

    435 GCCGTTCGCCATCTTCGT.........CAATGCGATGGACACCAATCCGC 475
        |||   ||||||   |||           |||      |||::::::::
    262 AlaGlyArgHisArgArgAlaHisAspProCysValProHisArgGlyCy 278

    476 TGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAA 525
        :|||   ||||||   ::::::|||:::::::::      |||   :::
    278 sGlyArgArgProCysAlaArgHisHisGlnGlyGlyGlyArgArgGluA 295

    526 CGCGGCCTGTT..................................... 536
          :::||||||
    295 rgHisProValLeuAspGlnSerHisHisAlaLeuHisArgGlnTyrGly 311

    537 ......................GGTATTGAGCCGCCTGACCGAACGTA 562
                               |||:::   ||||||:::|||
    312 GlyGlyAlaSerArgHisAlaAspGlyLeu...ProProSerArgProLe 327

    563 AAATCCATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCT 612
```

```
      :::|||                                    :::   |
  327 uAspPro.....................................ArgGlyC 332

  613 GCCAATATCGAAACACATGAATTTGGCGGCCCCGCATCCTGCCGGCTTGAG 662
      ||     |||::: :::           ||||||||||||      ::
  332 ysGlyLeuArgArg.Glu............ProHisProAlaArgLysTh 344

  663 TGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAACCGTGT 712
      :   :::        :::::::::   :::||||||          :
  344 rIleAlaAlaGluAspIleLeuHisAspMetGlyAla............P 357

  713 GGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACA 762
      ::::::::      :::|||      |||:::||||||:::
  357 heSerValIleSerSerAspSerGlnAlaMetGlyArgVal..... 370

  763 GGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAA 812
      |||::::::   |||:::         :::         ::::::|||::
  371 GlyGluValIleThrArgThrTrpGlnThrAlaHisLysMetLysValGl 387

  813 CAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTA 862
      ::::   ||||||   :::   :::   ||||||:::   :::|||   :
  387 nArgGlyArgLeuAlaGluGluThr...GlyAlaAsnAspAsnGlnArgV 403

  863 CCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTTCGGTA 912
      ::                      :::|||   |||:::     :::
  403 al.......................ArgArgTyrIleAlaLysTyrThr 411

  913 TTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGCTACCA 962
      :::|||   ||||||||||:::|||         ::::::|||
  412 IleAsnProAlaIleAlaHisGlyLeuSerArgHisIleGlySer..... 426

  963 CAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCT 1012
                  :::||||||||||:::      :::|||      |
  427 ..............ValGluGluGlyLysArgAlaAspLeuValLeuT 438

 1013 GGGTTGCGCCGCAGCCG 1029
      ||         ||||||
  438 rp.........GlnPro 440

seq_name: sp_invertebrate:Q9U0V2

seq_documentation_block:
ID   Q9U0V2      PRELIMINARY;      PRT;  2207 AA.
AC   Q9U0V2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   POSSIBLE MUS308 HOMOLOG (FRAGMENT).
GN   L8342.01.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RA   Masuy D., Purnelle B., Goffeau A., Ivens A.C., Lawson D., Quail M.,
RA   Rajandream M.A., Barrell B.G.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FRIEDLIN;
RX   MEDLINE=98146435; PubMed=9477341;
RA   Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA   Smith D.F.;
RT   "A physical map of the Leishmania major Friedlin genome.";
RL   Genome Res. 8:135-145(1998).
DR   EMBL; AL122012; CAB58415.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   SMART; SM00490; HELICc; 1.
KW   ATP-binding; Helicase.

FT   NON_TER    2207   2207
SQ   SEQUENCE   2207 AA;  229262 MW;  E461565DA0017B0C CRC64;


alignment_scores:
            Quality:  108.00              Length:    452
              Ratio:  0.565                Gaps:     23
 Percent Similarity:  42.257   Percent Identity: 22.788

alignment_block:
 US-09-303-518D-131/rev x Q9U0V2   ..

 Align seg 1/1  to: Q9U0V2  from: 1  to: 2207

 1232 CAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTAAATCGCGCAA 1183
      ||||||:::   |||   |||||||||:::   |||
  813 GlnProArgLeuCysGlnLeuSerValAlaThrThr.............. 824

 1182 AAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGC.....T 1145
             :::|||:::   |||   |||  :::|||          |
  825 .......ThrGlyGlnAlaSerAlaGlySerAlaArgAsnValAlaAlaS 839

 1144 CATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTC 1095
      ||   :::      ::::||   :::   |||||||||      |||
  839 erValAlaGluProAlaThrProValSerSerProProSerGluAla... 854

 1094 GTGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGT 1045

  854 ..................................................... 854

 1044 GATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGC 995
                          |||::::::|||:::|||      :::|||
  855 ..................CysAlaThrThrHisPro...ProAspLeuT 864

  994 TGCGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAA 945
          ||||||      ||||||
  864 rpLeuProSerLeuProThrGluHis..................... 872

  944 TCATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCAC 895
            ||||||                        :::|||   :::
  873 ValSerProProCys.....................AspProLeuLeuMe 882

  894 GCGGTTGTCCGCGTCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCG 845
       |||:::|||::::::   :::
  882 tArgValSerSerGlnGlyGly........................... 889

  844 CACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCC 795
                      |||   :::|||
  890 ................LysAlaGluGly..................... 893

  794 AAGGCAACCACGCGCTCGGTATTCAGACGGCCTGTTACGAACAAACGTCC 745
        ::::::::::::|||||||||||      |||:::        ||
  894 ...CysAlaAlaGlnSerValPheArgMetGlyValValAlaProThrPr 909

  744 GATAGCAATCACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGC 695
      |   ::::::   |||                        ||||||   :::
  909 oThrSerLeuGlySer.................AspValLeuSerSerA 920

  694 CGACTGGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGGCAGGATGC 645
        ||||||                        |||:::   |||::::::
  920 laThrGly..............ValSerAlaLeuSerAlaAlaAsnAla 931

  644 GGGCCGCCAAATTCATGTGTTTCGATATTGGCA.................. 612
      :::||||||   |||   :::   :::   |||
  932 AlaProProLeuSerAspLeuHisValThrAlaLeuProTyrLeuAlaTh 948

  611 .GCATTTTCAGACGGCACGTCTGCGCCTGCT.................... 582
       |||   :::   |||::::::|||||||||
  948 rAlaAlaAlaGlyGlySerGlyAlaProAlaValProAlaArgProGlyA 965

  581 ..........GCTTTACACACATGGATTTTACGTTCGGTCAGGCGGCTC 543
                 :::||||||:::            ::::::|||
  965 rgThrCysPheThrLeuHisSer............AlaAlaArgThrThr 977
```

```
    542 AATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTG..... 501
        :::|||           |||:::|||||||||||||:::
    978 GlyThr...........LeuSerSerSerAlaAlaAlaThrSerThrGl 990

    500 .......ATGATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATCG 458
                  :::::::::||||||:::::|||   |||      |
    990 uGluProThrThrAlaAlaAlaSerAlaSerThrGlyAlaValProProA 1007

    457 CATTGACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTG 408
        ||||||||:::   :::      ||||||:::   :::      ::::::
   1007 laLeuThrGluAlaSerTyrLeuSerAlaGluLysCys......PhePhe 1021

    407 AACGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCG 358
        :::|||   :::   :::      :::   |||
   1022 ArgGlyGlyThrAlaGluLeuGlnGlnLeuProGly.............. 1033

    357 GCGCACTTTTTCGCTGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTT 308
                        :::   ||||||   :::||||||      :
   1034 ................HisGlyAlaAlaSerProGlnGlyThr......G 1043

    307 CGAACTCGATTTCGTCGTTGCCTTCAACGGCAATCACGACTGACTGAAGT 258
        :::::   :::   ||||||   :::::|||   :::|||      :|:
   1043 luGlyLeuMetValSerLeuAspGlyValAlaSerAlaThrProSerAsp 1059

    257 ACGCGCTTTTCGCCACGGTGAATA........................... 234
        ||||||     |||      :::
   1060 ThrArgLysMetProSerSerMetGluSerArgHisThrThrGlyCysAl 1076

    233 .........................GCGGCGATTTTG.............. 222
                                  |||:::|||
   1076 aGlnProValIleAlaAsnAlaGlnProAlaLeuLeuArgProValAspS 1093

    221 .....CCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTTGTCT 177
             |||::::::   :::::::::   ||||||:::   :::   |||
   1093 erThrProSerThrValValMetHisArgThrProSerThrLeuThrSer 1109

    176 TCAAACAGCACTTGG..............CCTTTTTTGACGGCTTCACC 142
        ::::::   :::|||               ||||||           ||
   1110 GluGluProAlaTrpAsnArgArgSerAlaProPheProGlnProTrpPr 1126

    141 TTCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCA 92
        |||||||         :::|||:::   |||:::      :::   :::
   1126 oSerLeu......SerAspGlyGlnProProAlaLeuLeuThrValAlaA 1141

     91 ACGCGACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTG 42
          :::|||   :::::::::                  |||   |||
   1141 laValThrProLeuValGly.................CysMetGlyAla 1151

     41 CCCGCG 36
           |||
   1152 GluAla 1153
```

```
seq_name: sp_invertebrate:Q9VXM5

seq_documentation_block:
ID   Q9VXM5      PRELIMINARY;      PRT;  2406 AA.
AC   Q9VXM5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG9056 PROTEIN.
GN   CG9056.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003500; AAF48533.1; -.
DR   FlyBase; FBgn0030702; CG9056.
SQ   SEQUENCE   2406 AA;  248803 MW;  740EAD78D684F76C CRC64;
```

```
alignment_scores:
           Quality:  108.00            Length:    473
             Ratio:  0.519               Gaps:     28
 Percent Similarity:  43.975   Percent Identity: 22.833

alignment_block:
 US-09-303-518D-131/rev x Q9VXM5   ..

 Align seg 1/1  to: Q9VXM5  from: 1  to: 2406

   1298 CCGTAT.................TCGTATTTGCCCGGGCAGACGAAGCT 1267
        ||||||                 ||||||||||||           ::
   1266 ProTyrSerGlyLeuThrHisGlySerTyrLeuProProValLeuProVa 1282

   1266 G.................................CACAAAGCGAGGT 1253
        :                                 |||::::::   |
   1282 lAlaThrProAsnLeuSerAsnLeuProProThrGlnHisArgSer...S 1298

   1252 CTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCG 1203
        ||:::|||   ||||||::::::      |||
   1298 erAspSerArgAsnSerArgGluSerProAla.................. 1308

   1202 CCGACGATTAAATCGCGCAAAAGCAAGGTAGGC...AGGATGTCCAACGG 1156
           ::::::||||||      |||:::::|||   ::::::|||   ::
   1309 ...SerLeuLysSerThrProSerAsnIleGlyLeuAsnValSerMetAl 1324

   1155 CATTACG...CGCTCATAAGTGCCGATCGGT..............ACCA 1124
        :   |||   ||||||   :::|||::::::             :::
   1324 aProThrLeuArgSerIleThrProLeuAsnAsnSerSerAlaIleSerS 1341

   1123 TGGCGCGGTCGCCGCCGTTGACGGCTGTCGTG.........AACTTGAAG 1083
```

```
      :::   |||   |||:::::::::|||||          |||
1341 erGlyAlaSerGlnProValValSerValValProSerAlaAsnSerThr 1357

1082 AGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTG.....CGCGTGAT 1042
     :::|||   :::   :::   |||::::::            :::|||
1358 AlaLeuSerMetSerAsn...ProHisIleSerHisSerHisHisValPr 1373

1041 GGAGTATTTGTCCGGC.....TGCGGCGCAACCC 1013
       |||   ||||||          :::|||:::::
1373 oAlaTyrAlaSerGlyAlaPheSerSerSerAlaAlaAlaGlyThrSerT 1390

1012 AGCCGAACAGCTCTTTG......CTGCGGCCTTCTTCGATAACGGAAATC 969
     ||||||||||:::|||      |||      :::|||::::::
1390 hrProAsnSerGlyLeuSerThrLeuAlaValThrSerLeuSerThr... 1405

 968 TGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACC 919
                          |||:::||||||           ||
1406 .....SerAlaAlaPro.....GlnPr 1411

 918 GTTCAATACCGAACCGGAAATCACGCGG...TTGTCCGCGTCAACCAATT 872
     |   :::     |||:::   |||:::   |||     |||   |||
1411 oHisSerHisPheProGlnSerThrGlnMetLeuProGlnSerGlyAsnP 1428

 871 CGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGGTACGCAAGAGG 822
       :::|||||||
1428 heSerSerValSer..... 1432

 821 CGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATT 772
           |||||||||   :::|||   ::::::   ::::::   :::
1433 ...HisLeuLeuThrThrHisProMetSerSerGlnAsnGlnProMetVa 1448

 771 CAGACGGCCTGTTACGAACAAACGTCCGATAGCAATCACGTCTTGATAAT 722
      |||
1448 lArg..... 1449

 721 TGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATG 672
               :::::::|||::::::   :::::::::
1450 .....CysGlySerThrLeuTyrSerGlnSerSerAla..... 1460

 671 TGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGTT.. 624
                    |||:::   :::||||||:::::::::::::
1461 .................AlaAlaThrAlaProProSerAlaAlaAlaAl 1471

 623 .....TCGATATTGGCAG 611
                         |||:::   :::|
1471 aValSerAsnPheThrProSerValLeuAlaValGlnSerLeuThrThrA 1488

 610 CATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTA 561
     ||   :::::::::::::|||:::|||:::::::|||   :::   :::::
1488 laValThrSerSerSerSerSerProSerThrLeuSerSerSerValIle 1504

 560 CGTTCGGTC.....AGGCGGCTCAATACCAACAGGCC 529
     :::::::|||                    |||   :::|||
1505 GlnLysValIleSerProLysGlyGluSerProCysAsnLysAspArgAs 1521

 528 GCGTTTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAG 479
               ||||||   |||:::   ::::::|||:::   :::   :
1521 pSerSerTyrSerSerProAlaAsnAlaValValThrThrCysAlaProT 1538

 478 CCAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCA 429
     :::::   :::||||||                    :::||||||||||
1538 hrThrProIleValSer.....SerGlySerAla 1547

 428 TCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAA 379
           :::      |||   |||
1548 ArgProThrProProLeuSerAsn..... 1555

 378 GCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCA 329
                           |||         :::::::::::::
1556 .....CysThrSerMetGlyIleGlyMetValA 1565

 328 GCGCTTCAGGTACGTAGCGTTCG......AACTCGATTTCG.....TCG 291
     ::|||::::::|||   ||||||      |||:::|||||||      |||
```

```
1565 snAlaAlaSerThrAlaArgSerSerCysAsnAlaIleSerProLeuSer 1581

 290 TTGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACG 241
     :::|||:::|||
1582 IleProAlaThr..... 1585

 240 GTGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCG 191
           |||:::|||      :::|||          |||:::|||:
1586 .......AlaGlyIleHisValSerAla.....ThrAsnProS 1596

 190 GATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCT 141
     ::|||      :::|||:::|||   :::|||          |||   |||
1596 erPhe......GlnSerSerSerTyrPheProThrProLeuAlaProPro 1610

 140 TCCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAA 91
                                              ||||||||||||
1611 Pro.....SerSerProSerPr 1616

  90 CGCGACTTCGGTAATGGCC 72
     |||||||||      |||
1616 oAlaThrSerSerAlaAla 1622
```

seq_name: sp_human:Q9H195

```
seq_documentation_block:
ID   Q9H195      PRELIMINARY;      PRT;   901 AA.
AC   Q9H195;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MUC3B MUCIN (FRAGMENT).
GN   MUC3B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pratt W.S., Crawley S., Hicks J., Ho J., Nash M., Kim Y.S., Gum J.R.,
RA   Swallow D.M.;
RT   "Multiple transcripts of MUC3: Evidence for two genes, MUC3A and
RT   MUC3B.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ291390; CAC19572.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000082; SEA.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS50024; SEA; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   901 AA;  96192 MW;  F98FC817494ECD99 CRC64;
```

```
alignment_scores:
            Quality:  107.00            Length:    440
              Ratio:  0.532               Gaps:     16
 Percent Similarity:  45.682   Percent Identity: 20.909

alignment_block:
 US-09-303-518D-131/rev x Q9H195   ..

 Align seg 1/1  to: Q9H195  from: 1  to: 901

    1292 TCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTC 1243
         |||   :::      :::|||      |||::::::   |||   |||||
       6 SerIleIleThrThrGluThrThrSerHisSerThrProSerPheSerSe 22

    1242 CAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATTA 1193
         |::::::   :::                  :::||||||   :::
      22 rSerThrIleHisSer.....ThrValSerSerSerThrT 34

    1192 AATCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTCA 1143
```

```
      :::      |||                 ::::::|||:::|||   :::
   34 hrAlaIleThrSerProPheThrThrAlaGluThrGlyValThrSerThr 50

 1142 TAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGT 1093
                :::::::::::::   :::   |||   |||:::
   51 ProSerSerProSerSerLeuSerThrAspIleProThrThrSer..... 65

 1092 GAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGA 1043
                                                :::|||:::::
   66 .....LeuArgThrL 69

 1042 TGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTG 993
      ::      ||||||          :::|||:::   :::::::::
   69 euThrProLeuSerLeuSerThrSerThrSerLeuThrThrThrThrAsp 85

  992 CGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATC 943
        ||||||      |||:::|||                          ||
   86 LeuProSerIleProThrAspIle.....Se 94

  942 ATGCGCGCCTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGC 893
      |      |||      |||      :::::::::::::|||:::|||
   94 rSerLeuProThrProIleHisIleIleSerSerSerProSerIleGlnS 111

  892 GGTTGTCCGCGTCA.....ACCAATTCGCCGGCGGTAAGTTGA 855
           ::::::|||           |||:::|||||:::::::|||
  111 erThrGluThrSerSerLeuValGlyThrThrSerProThrMetSer... 126

  854 GACACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTG 805
                      |||||||||
  127 ................ThrValArg..... 129

  804 CAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGTTACGA 755
                |||||   ||||||:::   :::   |||:::::::
  130 ............AlaThrLeuArgSerThrGluAsnThrProIleSerS 142

  754 AC...AAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACACG 708
      ::   :::      |||:::::::|||
  142 erPheSerThrSerIleValValThrPro..... 151

  707 GTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCACTCAA 658
            :::||||||                 ::::::   |||:::
  152 ......GluThrProThrThr...........GlnAlaProProValLe 162

  657 GCCGGCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGCAT 608
          ::::::   |||      :::|||   |||         ::::::|
  162 uMetSerAlaThrGlyThrGlnThrSerProValProThrThrValThrP 179

  607 TTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGT 558
      ||::::::      |||:::   ::::::|||||||||
  179 heGlySerMetAspSerSerThrSerThrLeuHisThrLeuThrProSer 195

  557 TCGGTC......AGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTC 514
      ::::::      :::   ::::::|||::::::      :::   |||::
  196 ThrAlaLeuSerLysIleMetSerThrSerGlnPheProIleProSerTh 212

  513 GGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGT 464
      :   :::|||   ::::::|||:::   |||      |||      :::|
  212 rHisSerSerThrLeuGlnThrThrProSerIleProSerLeuGlnThrS 229

  463 CCATCGCATTGACG.................AAGATGGCGAACGGCTCG 432
      ||::::::   |||                :::::::::::|||||||
  229 erLeuThrSerThrSerGluPheThrThrGluSerPheThrArgGlySer 245

  431 GCATCTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCA 382
      :::||||||   :::||||||
  246 ThrSerThrAsnAlaIleLeu..... 252

  381 TAAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTG 332
                              :::||||||   :::   ::::
  253 ......................ThrSerPheSerThrIleIleTrpS 261

  331 CCAGCGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCTTCA 282
      ::|||:::               :::      :::||||||   ||||||
```

```
  261 erSerThrPro...........ThrIleIleMetSerSerSerProSer 273

  281 ACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGC 232
      :::|||      |||
  274 SerAlaSerIleThr.............................Pr 279

  231 GGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTT 182
       ::::::               :::|||   :::::::|||   :::|||
  279 oValPheAlaThrThrIleHisSerValProSerSerProTyrIlePheS 296

  181 TGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTTCCTTGATT 132
        :::::::|||               :::||||||   |||||||||
  296 erThrGluAsnValGlySerAlaSerIleThrAlaPheProSerLeu... 311

  131 TTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAACGCGACT.. 84
            ::::::         |||   :::||||||::::::::::::
  312 ......SerSerSerSerThrThrSerThrSerProThrSerSerSerLe 326

   83 ..............................TCGGTAATGGCCGGGCCGT 65
                                    |||   ::::::   ||||
  326 uThrThrAlaLeuThrGluIleThrProPheSerTyrIleSerLeuProS 343

   64 CATAAATGACTTGCTCCGGT 45
      ||         |||   |||
  343 erThrThrProCysProGly 349
```

seq_name: sp_rodent:Q9QY35

seq_documentation_block:
ID Q9QY35 PRELIMINARY; PRT; 1979 AA.
AC Q9QY35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HOST CELL FACTOR C1 (FRAGMENT).
GN ABCD1 OR HCFC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Platzer M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
RA Rosenthal A.;
RT "Comparative sequence analysis of the mouse Llcam locus and the
RT corresponding region of human Xq28.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF133093; AAF22156.1; -.
DR MGD; MGI:1349215; Abcd1.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SM00060; FN3; 1.
FT NON_TER 1 1
SQ SEQUENCE 1979 AA; 203436 MW; ABE998C5DDE688A9 CRC64;

alignment_scores:
Quality: 107.00 Length: 464
Ratio: 0.510 Gaps: 24
Percent Similarity: 45.259 Percent Identity: 22.198

alignment_block:
US-09-303-518D-131/rev x Q9QY35 ..

Align seg 1/1 to: Q9QY35 from: 1 to: 1979

```
 1118 CGGTCGCCGCCGTTGACGGCTGTC.....GTGAACTTGAAGAG 1081
      :::   ||||||   :::   |||          :::|||   :::||
  297 LysProProProProAlaArgValGlnLeuValArgAlaAsnThrAsnSe 313

 1080 TTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTG. 1032
```

```
313 rLeuGluVal...SerTrpGlyAlaValAlaThrAlaAspSerTyrLeuL 329

1031 ..........................TCCGGCTGCGGCGCAACCCAG 1011
                               :::::::::   ||||||:::
329 euGlnLeuGlnLysTyrAspIleProAlaThrAlaAlaThrAlaThrSer 345

1010 CCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTG 961
     |||:::               |||:::   :::   ::::::
346 ProThr..............ProAsnProValProSerValProAla.. 356

960 GTAGCGTCCCAAATAATCATGCGCGCCTTGTGCAATCGCACCG....... 918
    :::|||     |||   ||||||:::|||   ||||||
357 ....AsnProProLysSerProAlaProAlaAlaAlaAlaProAlaValG 372

917 ...................TTCAATACCGAACCGGAAATCACGCGGTTG 888
                        ::::::      |||:::   :::
372 lnProLeuThrGlnValGlyIleThrLeuValProGlnAlaAlaThrAla 388

887 TCCGCGTCAACCAATTCG............CCGGCGGTAAGTTGAGACAC 850
          ||||||::::::             |||:::|||      :::::
389 ProProSerThrThrThrIleGlnValLeuProThrValProGlySerSe 405

849 CTTCGCACCCAAAACGGTACGCAAGAGGCGCGGT.........TTGTTGA 809
    :::::::      |||:::   :::   :::|||          :::|||
405 rIleSerValProThrAlaAlaArgThrGlnGlyValProAlaValLeuL 422

808 CTTGCAGGCCGCCCAAGGCAACCACGCGCTCG......GTATTCAGACGG 765
              |||:::|||||||||   :::       |||      |||
422 ysValThrGlyProGlnAlaThrThrGlyThrProLeuValThrMetArg 438

764 CCTGTTACG......AACAAACGTCCGATAGCAATCACGTCTTGATAATT 721
    |||::::::      :::|||   |||:::::::::|||||      ||
439 ProAlaSerGlnAlaGlyLysAlaProValThrValThrSer......Le 453

720 GATGGTCCACACGGTTTTATTCGCGCCGACTGGCTCGATGAAATGAATGT 671
    |   :::   :::   :::   :::|||||   |||   :::
453 uProAlaSerValArgMetValValProThrGlnSerAlaGlnGlyThrV 470

670 GCGTGCCACTCAAGCCGGCAGGATGCGGG..................... 642
      :::      :::|||         |||
470 alIleGlySerAsnProGlnMetSerGlyMetAlaAlaLeuAlaAlaAla 486

641 ....................CCGCCAAATTCATGTGTTTCGATATTGGC 613
                        ||||||:::|||:::   :::::::|||::
487 AlaAlaAlaThrGlnLysIleProProSerSerAlaProThrValLeuSe 503

612 A.............................GCATTTTCAGACGGCA 596
    :                            |||   :::   ||||
503 rValProAlaGlyThrThrIleValLysThrValAlaValThrProGlyT 520

595 CGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGTCAGGCGG 546
    ||::::::      |||                    |||
520 hrThrThrLeuProAlaThrValLysValAlaSerSerProValMetVal 536

545 CTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTTTGATGAT 496
       |||              ||||||:::::|||||||::::::
537 SerAsnProAlaThrArgMetLeuLysThrAlaAlaAlaGlnValGlyTh 553

495 GACCGTAGGGTCGGCAGCC.............................. 477
     :::|||:::||||||||||
553 rSerValSerSerAlaAlaAsnThrSerThrArgProIleIleThrValH 570

476 .....AGCGGATTGGTGTCCATCGCA..............TTGACGAAG 447
         ||||||   |||::::::|||                ::::|||
570 isLysSerGlyThrValThrValAlaGlnGlnAlaGlnValValThrThr 586

446 ATGGCGAACGGCTCGGCATCTACG.......................... 423
    :::::::::|||   ::::::|||
587 ValValGlyGlyValThrLysThrIleThrLeuValLysSerProIleSe 603

422 ............GCAGGGATTTTGCTGAAC...GGACGGGTGCGAAGCG 389
                :::::::::::::   |||   |||:::|||   ||||:

603 rValProGlyGlySerAlaLeuIleSerAsnLeuGlyLysValMetSerV 620

388 CAGTCCAT...AAGCCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTG 342
    ::|||:::   ||||||                      :::|||   ::::::
620 alValGlnThrLysPro..............ValGlnThrSerAlaVal 631

341 CTCAATTTTGCCAGCGCT.............................. 324
       :::   ||||||:::
632 ThrGlyGlnAlaSerThrGlyProValThrGlnIleIleGlnThrLysGl 648

323 ..........TCAGGTACGTAGCGTTCG......AACTCGATTTCGTCGT 290
                ::::|||||||      :::      :::|||   ::::::
648 yProLeuProAlaGlyThrIleLeuLysLeuValThrSerAlaAspGlyL 665

289 TGCCTTCAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGG 240
      |||:::|||   |||||||||
665 ysProThrThrIleIleThrThr........................... 672

239 TGAATAGCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGG 190
                              :::|||:::::::::::::|||   |||
673 ..................ThrGlnAlaSerGlyAlaGlyThrLysPro.. 682

189 ATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTT 140
       :::|||:::   :::|||:::   |||      |||        :
683 .ThrIleLeuGlyIleSerSerValSerProSerThrThrLysProGlyT 699

139 CCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAAC 90
    ::      :::|||:::                    :::   |||   :::
699 hrThrThrIleIleLys.................ThrIleProMetSer 709

 89 GCGACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCC 48
    |||      ::::::||||||   :::   ::::|||   |||
710 AlaIleIleThrGlnAlaGlyAlaThrGlyValThrSerSer 723
```

seq_name: sp_human:Q14395

seq_documentation_block:

```
ID   Q14395      PRELIMINARY;      PRT;   505 AA.
AC   Q14395;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MUCIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=STOMACH;
RX   MEDLINE=93194895; PubMed=7680650;
RA   Toribara N.W., Roberton A.M., Ho S.B., Kuo W.L., Gum E., Hicks J.W.,
RA   Gum J.R. Jr., Byrd J.C., Siddiki B., Kim Y.S.;
RT   "Human gastric mucin. Identification of a unique species by expression
RT   cloning.";
RL   J. Biol. Chem. 268:5879-5885(1993).
DR   EMBL; L07518; AAB61945.1; -.
FT   NON_TER       1      1
FT   NON_TER     505    505
SQ   SEQUENCE   505 AA;  51967 MW;  45D7CB1E6050E07D CRC64;
```

alignment_scores:

```
            Quality:  106.50              Length:    492
              Ratio:  0.475                 Gaps:     21
 Percent Similarity:  45.528   Percent Identity: 20.325
```

alignment_block:

US-09-303-518D-131/rev x Q14395 ..

Align seg 1/1 to: Q14395 from: 1 to: 505

```
   1340 CCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTATTC 1291
```

```
    |||:::   :::   :::|||:::|||      :::   :::|||   ||
 85 ProGluValAlaProThrSerThrThrThrIleThrProAsnProThrSe 101

1290 GTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCA 1241
     |          :::|||:::          |||:::::::::|||::::
 101 r.........ThrArgThrArgThrProValAlaHisThrAsnSerAlaT 115

1240 ATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCG...CCGACGATT 1194
     ::|||:::::|||              :::   |||   ||||||
 115 hrSerSerArgProProProProPheThrThrHisSerProProThrGly 131

1193 AAATCGCGCAAAAGCAAGGTAGGCAGGATG......TCCAACGGCATTAC 1150
     :::|||      |||:::::|||   |||      ::::::      ||
 132 SerSerProPheSerSerThrGlyProMetThrAlaThrSerPheLysTh 148

1149 GCGCTCATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGG 1100
     |   :::     |||      ::::::   ::::::   |||||||||
 148 rThrThrThrTyrProThrProSerLeuProGlnThrThrProLeuThrH 165

1099 CTGTCGTGAACTTGAAG.............................. 1083
       |||      ::::::
 165 isValProProPheSerThrSerLeuValThrProIleThrHisThrVal 181

1082 ............................AGTTTGTTTTTTAGGAAATG 1063
                                 |||   ::::::::::::
 182 IleThrProThrHisProGlnMetSerThrSerAlaTyrIleHisSerTh 198

1062 GCCG.......................................... 1059
      |||
 198 rProThrGlyThrIleAlaSerProThrThrValLysAlaThrArgSerT 215

1058 .........................AGAGTGGTGCGCGTGATGGAGTAT 1035
                             :::   |||:::   :::
 215 hrTyrThrAlaProLeuMetThrAlaThrThrSerArgIleSerGlnAla 231

1034 TTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTC 985
       |||:::   ::::::::::::   ::::::||||||        ::
 232 HisSerSerIleSerThrAlaLysThrSerThrSerLeuHisSerHisAl 248

 984 TTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGC 935
     :|||   |||:::
 248 aSerSerThrHis..................................... 252

 934 CTTGTGCAATCGCACCGTTCAATACCGAACCGGAAATCACGCGGTTGTCC 885
                              :::||||||:::|||      |||
 253 ........................HisProGluValThrProThrSer 260

 884 GCGTCAACCAATTCGCCG...GCGGTAAGTTGAGACACCTTCGCACCCAA 838
     :::::::::   :::|||   :::::::|||   ||||||   :::|||
 261 ThrThrAsnValThrProLysSerThrSerArgAspThrSerThrPro.. 276

 837 AACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCC....... 795
         |||          :::      :::   |||||||||
 277 ....ValThrHisThrThrSerAlaThrSerSerArgProProThrProI 292

 794 ...........AAGGCAACCACGCGCTCGGTATTCAGACGGCCTGTTACG 756
                :::::   |||||||||           |||::::::
 292 leThrThrHisSerSerProThrArgSer.........SerProLeuSer 305

 755 AACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACACGGT 706
     :::      |||::::::   ||||||              :::   |||::
 306 SerThrGlyProMetThrAlaThrSer............IleLysThrTh 318

 705 TTTATTCGCGCCGACTGGCTCGATGAAATGAATGTGCGTGCCACTCAAGC 656
     :        ||||||   |||
 318 rThrThrTyrProThrProSer........................... 325

 655 CGGCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGCATTT 606
                  |||   ::::::   ::::::   :::      |||
 326 ...........HisProGlnThrThrLeuThrThrHisValProProPhe 338

 605 TCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTC 556
     |||   ::::::   :::|||::::::   ||||||   |||      ::

 339 SerThrSerSerValThrProSerThr...HisThrValIleThrProTh 354

 555 GGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTT 506
     :      :::::::::|||:::
 354 rHisAlaGlnMetSerThrSer.......................... 361

 505 CTTTGATGATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCA 456
                 ::::::   |||:::   :::|||   |||
 362 ........AlaSerIleHisSerThrProThrGlyThrValProProLeu 375

 455 TTGACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGAA 406
       |||:::|||   :::||||||:::   |||   |||      :::::
 376 ThrThrArgMetProThrGlySerThrArgThr...GlyProProMetTh 391

 405 CGGACGGGTGCGAAGCGCAGTCCATAAGCCTGATTGAATCAGGTTGCGGC 356
     :|||   :::                                |||:::      :
 391 rGlyThrIle.........................IleGlnThrSerL 399

 355 GCACTTTTTCGCTGCTCAATTTTGCCAGCGCTTCAGGTACGTAGCGTTCG 306
     :::::   :::   ::::::   |||:::::::|||   ::: - :::|||
 399 ysAlaHisAsnSerPheSerThrAlaLysThrSerThrSerLeuHisSer 415

 305 AACTCGATTTCGTCG...TTGCCTTCAACGGCAATCACGACTGACTGAAG 259
     ::::::   |||:::      |||:::|||:::   |||:::      ::
 416 HisAlaSerSerThrHisHisProGluThrThrProThrSer......Th 430

 258 TACGCGCTTTTCGCCACGGTGAATAGCGGCGATTTTGCCTGAAGCCGGCG 209
     :|||::::::::::|||:::                          :::|||||||:
 430 rThrAsnIleThrProLys..............SerThrSerAlaGlyT 442

 208 CAGTAAATACTACGCCCGGATTCTTTTTGTCTTCAAACAGCACTTGGCCT 159
     ::   :::   :::            |||::::::::::|||      |||
 442 hrSerThrProValAlaHisThrThrLeuAlaThrSerSerArgLeuPro 458

 158 ......TTTTTGACGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGCAT 115
           |||   |||      |||
 459 ThrThrPheThrThrHisPhePro........................... 466

 114 GCCGACATATTCTTCGCCAAGCAACGCGACTTCGGTAATGGCCGGGCCGT 65
      ||||||   ||||||      :::::::|||:::   |||::::::   |
 467 .ProThrGlySerSerHisValSerSerThrGlyProMetThrAlaThrS 483

  64 CATAAATGACTTGCTCCGGTCTGCCC 39
     ||   :::|||   :::      |||
 483 erSerGlnThrThrThrThrHisPro 491
```

```
seq_name: sp_bacteriap:Q98K15

seq_documentation_block:
ID   Q98K15      PRELIMINARY;      PRT;   515 AA.
AC   Q98K15;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   ATP-BINDING PROTEIN OF RIBOSE ABC TRANSPORTER.
GN   MLL1680.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP002998; BAB48999.1; -.
```

```
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00005; ABC_tran; 2.
DR   SMART; SM00382; AAA; 1.
KW   ATP-binding; Complete proteome.
SQ   SEQUENCE   515 AA;  55185 MW;  882AC3E521F5A1B6 CRC64;


alignment_scores:
           Quality:  106.50              Length:    483
             Ratio:  0.498                Gaps:     17
 Percent Similarity:  44.306   Percent Identity: 18.841

alignment_block:
 US-09-303-518D-131 x Q98K15   ..

 Align seg 1/1  to: Q98K15  from: 1  to: 515

      88 GCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTCGATGAAAATCAA 137
         |||:::|||||||||:::   :::|||          :::|||:::
      37 AlaIleLeuGlyGluAsnGlyAlaGlyLysSerThrLeuIleLysValMe 53

     138 GGAAGGT...GAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAA 184
            |||      |||      :::|||:::::::   ::::::   ::::
      53 tAlaGlyValValAlaProAspGluGlyArgMetThrLeuAspGlyArgG 70

     185 AGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAATCGCCGCT 234
         ::         |||:::|||   :::||||||:::::::::   |||:::
      70 lu.........ValThrPheLeuSerProAlaAlaAlaAsnGlnAlaGly 83

     235 ATT.......................................... 237
         |||
      84 IleValCysIlePheGlnGluLeuSerLeuValProAspLeuSerValAl 100

     238 ..............................................C 238
                                                      :
     100 aAspAsnIleValIleSerAspProProLysArgPheGlyMetIleAspA 117

     239 ACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATTGCCGTTGAAGGC 288
         ::::::::::::::||||||   ::::::::::::      :::   :::
     117 rgLysAlaGlnArgArgValAlaGluGluAlaLeuAlaArgAlaGlyAla 133

     289 AACGAC.......................GAAATCGAGTTCGAACGCTA 314
            |||                      ::::::   ::::::|||
     134 AlaAspIleHisProLeuAlaLeuValLysAspLeuProLeuSerArgAr 150

     315 CGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACC 364
                 :::|||||||      ::::::|||   |||      :
     150 gGlnMetValGluIleAlaLysAlaLeuAlaArgLysProArgIleLeuI 167

     365 TGATTCAA.........TCAGGCTTATGGACTGCGCTTCGCACCCGTCCG 405
         ::::::::         |||:::|||   :::|||      ::::::
     167 leLeuAspGluAlaThrSerAlaLeuThrAlaAlaAspValSerLysIle 183

     406 TTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCGTCAA 455
         |||:::   :::      :::  :::|||   :::|||::::::
     184 PheGlyValLeuLysArgLeuArgSerGluGlyLeuAlaLeuLeuTyrIl 200

     456 TGCGATGGACACCAATCCGCTGGCT......GCCGACCCTACGGTCATCA 499
          ::       |||   :::|||      ||||||      :::::::
     200 eSerHisArgMetAsnGluIleAlaGluLeuAlaAspGlnCysThrValP 217

     500 TCAAAGAAGCC......GCCGAAGACTTCAAACGCGGCCTGTTGGTATTG 543
         ::::::::::      :::   :::::|||   |||
     217 heArgAsnGlyArgAsnValAlaSerTyrLysAlaGlySerLysSerAsp 233

     544 AGCCGCCTGACCGAACGTAAAATC..............CATGTGTGTAA 578
         ::::::::::::|||      |||              |||:::
     234 AsnGluValValGluLeuMetIleGlyArgGluTyrSerHisIlePhePr 250

     579 AGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTGCCAATATCGAAACAC 628
          |||      |||   |||||||:::      |||   ::::::|||

     250 oAlaLysProAlaAlaValProAlaThrAlaAlaProArgLeuGlu.... 265

     629 ATGAATTTGGCGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCAT 678

     265 ..................................................... 265

     679 TTCATCGAGCCAGTCGGCGCGAATAAAACCGTGTGGACCATCAATTATCA 728
                            |||:::|||      ||||||   :::   ::
     266 .................AlaArgLysLeuSerTrpThrAspArgLeuHi 276

     729 AGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCGTCTGAATACCG 778
         :|||:::            |||                        :::
     276 sAspIleSerLeuThrValArg.......................AlaG 285

     779 AGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAACCGCGCCTCTTG 828
           :::||||||:::|||||||||||||:::   :::      |||   |||
     285 lyGluValValGlyLeuGlyGlyLeuAspGlyGlnGlyGlnArgGluLeu 301

     829 CGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCCGGCGAATTGGT 878
               ::::::   :::   :::      |||::::::      |||::
     302 LeuLeuAlaPhePheGlyValLeuArgGlyLeuSerGlyGluValLeuIl 318

     879 TGACGCGGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGTGCGATTG 928
         :|||:::                        |||      |||:::  |
     318 eAspGlyLys.................ProValAlaIleGlySerProA 329

     929 CACAAGGCGCGCATGATTATTTGGGACGCTACCACAATCAGATTTCCGTT 978
         ||::::::   :::|||   :::|||                 :::::::::
     329 laLysAlaArgGlnAspGlyIleGly..............MetAlaLeu 340

     979 ATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCC 1028
         |||   |||   |||:::   |||      |||   :::   |||:::
     341 IleProGluAspArgLysThrGlu......GlyLeuMetLeuProMetTh 355

    1029 GGACAAATACTCCATCACGCGCACCACTCTCGGCCATTTCCTAAAAAACA 1078
            :::   :::::::::   ::::::|||   ::::::   |||:::
     355 rValArgGluAsnLeuSerPheAlaAlaLeuAspArgLeuSerLysGlyG 372

    1079 AACTCTTCAAGTTCACGACA............................. 1098
            ::::::      ::::::
     372 lyIleIleAspArgAlaAlaGluGlnArgLeuIleAspAspMetValGly 388

    1099 .....................................GCCGTCAA 1106
                                              |||::::::
     389 LeuLeuAlaIleLysThrAlaGlyLeuAspIleProValGlyAlaLeuSe 405

    1107 CGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGAGCGCGTAATGC 1156
         :||||||::::::      |||   |||:::   :::      :::|
     405 rGlyGlyAsnGlnGlnLysValValIleAlaLysTrpLeuMetArgGlnP 422

    1157 CGTTGGACATCCTG.........CCTACC..................... 1176
         ||       ||||||         ||||||
     422 roArgIleIleLeuLeuAsnAspProThrArgGlyIleAspValGlyThr 438

    1177 ..............TTGCTTTTGCGCGATTTA................. 1194
                           |||:::|||   |||
     439 LysGlnGluLeuTyrGlnLeuMetArgLysLeuAlaAspAlaGlyAlaAl 455

    1195 ....ATCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAA 1239
            :::      :::||||||      :::   :::||||||   :::
     455 aIleLeuPheTyrSerThrAspTyrAspGluLeuIleGlyCysCysAsp 471

seq_name: sp_bacteria:Q46767

seq_documentation_block:
ID   Q46767      PRELIMINARY;      PRT;   732 AA.
AC   Q46767;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF732.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

```
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=49;
RX   MEDLINE=96038471; PubMed=7480171;
RA   Lim D.;
RT   "Analysis of a retron EC86 and EC67 insertion site in Escherichia
RT   coli.";
RL   Plasmid 34:58-61(1995).
DR   EMBL; U23723; AAA64865.1; -.
SQ   SEQUENCE   732 AA;  81092 MW;  D4576038D605D6D8 CRC64;


alignment_scores:
            Quality:  106.50              Length:    466
              Ratio:  0.475                Gaps:     25
 Percent Similarity:  48.069   Percent Identity: 22.747

alignment_block:
 US-09-303-518D-131 x Q46767   ..

 Align seg 1/1  to: Q46767  from: 1  to: 732

      28 CTGCCCATCGCGGGCAGACCGGAGCAAGTCATTTATGACGGCCCGGCCAT 77
         ||||||:::   :::   ||||||              |||   |||::
      57 LeuProLeuLeuAlaMetProGlu..............GlyLysAlaVa 68

      78 TACCGAAGTC............GCGTTGCTTGGCGAAGAATAT....... 108
         :::::::               |||:::|||::::::::::|||
      68 lAlaLysPheValLeuHisProAlaPheLeuAlaLysSerTyrPheProV 85

     109 ............................GTCGGCATGCGCCCCTCG 126
                                     :::|||   :::   :::
      85 alSerLeuPheGluArgPheSerLeuGluSerIleGlySerLys...Ala 100

     127 ATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTGA 176
         :::|||:::|||      :::   :::||||||              ::
     101 ValLysValLysProArgLysAspIleLysLys..............Ar 112

     177 AGACAAAAAGAATCCGGGCGTAGTATTTACTGCGCCGGCTTCAGGCAAAA 226
         :   :::|||:::       ::::::           :::||||||||||
     112 gGlyArgLysGluGluTyrThrThrAlaCysIleTyrValSerGlyLysL 129

     227 TCGCCGCTATTCACCGTGGCGAAAAGCGCGTACTTCAGTCAGTCGTGATT 276
               ::::::                  |||   |||::::::::::
     129 ysGluSerProSerPheLeuAspSerValAsnGlnAspThrLeuThr 145

     277 GCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTAC.......... 315
               :::   ||||||   :::   :::|||:::
     146 LysGlyGlnGlnAsnAspPheValThrLeuGluAsnIleSerIleLeuGl 162

     316 .GTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGCAACC 364
          |||   :::   :::      ::::::|||::::::|||
     162 uValSerAspLysValLysThrIleAsnSerAsnGluVal.......... 175

     365 TGATTCAATCAGGCTTATGGACTGCGCTTCGCACCCGTCCGTTCAGCAAA 414
                    ::::::   :::||||||:::|||   |||   ::::::
     176 ........MetSerIleGluValAlaLeuHisThr...ProAspThrSer 188

     415 ATCCCTGCCGTAGATGCC...GAGCCGTTCGCC.................. 444
                 ||||||:::   |||   ||||||
     189 SerSerIleValAspSerPheGluValPheAlaSerGlnAsnGlyAlaVa 205

     445 ...................ATCTTCGTCAATGCGATGGACACCAATCCGC 475
                           |||   |||:::::::::         |||:
     205 lIleAspLysAlaArgSerIleLysValLysGlyLeuThrPheMetProI 222

     476 TGGCTGCCGACCCTACGGTCATCATCAAAGAAGCCGCCGAAGACTTCAAA 525
         ::   |||:::      |||   :::|||   |||      :::|||
     222 leLysAlaSerLysAspValAlaLeuLysValAlaGluPheSerPheLeu 238

     526 CGCGGCCTGTTGGTATTGAGC......CGCCTGACCGAACGTAAAATC.. 567

         |||   |||      |||         ||||||:::|||      |||
     239 ArgThrLeuArgGluLeuProGluLeuArgLeuSerGluProValIleSe 255

     568 ....CATGTGTGTAAAGCAGCAGGCGCAGACGTGCCGTCTGAAAATGCTG 613
                |||   ::::::::::::   ::::::|||||||||:::|||:
     255 rArgSerValIleGlnThrSerAsnLeuSerLeuProSerGluGlyAlaV 272

     614 CCAAT......ATCGAAACACATGAATTTGGCGGCCCGCATCCTGCCGGC 657
         ::|||      |||::::::      |||   |||
     272 alAsnProHisIleLysValAlaIlePheAspGly.............. 283

     658 TTGAGTGGCACGCACATTCATTTCATCGAGCCAGTCGGCGCGAATAAAAC 707
               |||      |||      :::::::|||
     284 ......GlyLeuGlyIleAspAspPheAsnPro................. 292

     708 CGTGTGGACCATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCG 757
              |||:::   :::|||
     293 ....TrpValThrGluTyrThr.........................P 299

     758 TAACAGGCCGTCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAA 807
           :::|||:::   :::|||:::      :::         |||   :::
     299 heAsnGlyAsnAlaSerThrAsnAlaLysLeuLeuSerHisGlyGlnAsp 315

     808 GTCAACAAACCGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCA 857
         |||::::::      :::|||      |||||||||:::::::::::::
     316 ValThrSer...ThrValLeuPheGlyValLeuGlySerGluThrGluLy 331

     858 ACTTACCGCCGGCGAATTGGTTGACGCGGACAACCGCGTGATTTCCGGTT 907
         :|||::::::                          |||||:::::::::::
     331 sLeuSerValProTyrCysAsnIleAspHisTyrArgValLeuAspSerA 348

     908 CGGTATTGAACGGTGCGATTGCACAAGGCGCGCATGATTATTTGGGACGC 957
         ::|||   |||:::   :::                 |||   |||   |||
     348 snValAsnAsnSerAspVal......AspLeuPheAspValLeuIleArg 362

     958 TACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
                         ||||||:::|||:::   :::         :::
     363 IleLys.........SerValLeuGluGlnLysLysTyrAspTyrIleAs 376

    1008 CGGCTGGGTTGCGCCGCAG............CCGGACAAATACTCCATCA 1045
                  ::::::|||:::                |||         :::
     376 nLeuSerLeuGlyProArgLeuProValAspAspAspAspValHisValT 393

    1046 CGCGCACCACTCTCGGCCATTTCCTAAAAAACAAACTCTTCAAGTTCACG 1095
              :::||||||   :::::::|||   :::                |||
     393 rpThrSerThrLeuGluGluIleLeuAlaSerGlyGluThrLeuCysThr 409

    1096 ACAGCCGTCAACGGCGGCGACCGCGCCATGGTACCGATCGGCACTTATGA 1145
         :::||||||      |||:::   ::::::::|||   :::      ::
     410 ValAlaVal......GlyAsnAspGlyGlnLeuPro...AlaLysLeuAs 423

    1146 GCGCGTAATGCCGTTGGACATCCTGCCTACCTTGCTTTTGCGCGATTTAA 1195
         :|||::::::|||            |||:::   |||::::::   |||
     423 nArgIleGlnPro............ProAlaAspLeuValAsnGlyLeuS 436

    1196 TCGTCGGCGATACCGACAGCGCGCAGGCTTTGGGTTGCTTGGAATTGGAC 1245
           ||||||   :::   |||
     436 erValGlyAlaAlaThrSerLeu.......................... 443

    1246 GAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGC......CCGGGCAAA 1287
         :::::::::         ||||||:::   |||      ||||||:::
     444 SerAspSerTrpGluArgCysSerTyrSerCysIleGlyProGlyArg 459

seq_name: sp_bacteriap:O51099

seq_documentation_block:
ID   O51099      PRELIMINARY;      PRT;   773 AA.
AC   O51099;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 89.8 KDA PROTEIN.
GN   BB0072.
```

```
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
DR   EMBL; AE001120; AAC66467.1; -.
DR   TIGR; BB0072; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   773 AA;  89847 MW;  42A49219431EF362 CRC64;


alignment_scores:
           Quality:  106.50              Length:    305
             Ratio:  0.678                Gaps:     17
 Percent Similarity:  51.475   Percent Identity: 22.623

alignment_block:
 US-09-303-518D-131 x O51099   ..

 Align seg 1/1  to: O51099  from: 1  to: 773

      76 ATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGTCGGCATGCGCCCCTC 125
         |||   |||   ||||||:::      |||      :::   :::
      26 IleProGluTyrAlaLeuIleProLeuGluThrGluAsnSerLysSerIl 42

     126 GATGAAAATCAAGGAAGGTGAAGCCGTCAAAAAAGGCCAAGTGCTGTTTG 175
          :::   |||   |||   :::   :::::::::::::|||:::|||
      42 eIleTyrIleIleGluLysGlnLysIleGluGluAsnGlnIleLeu.... 57

     176 AAGACAAAAAAGAATCCGGGCGTA...GTATTTACT...GCGCCGGCTTCA 219
               :::||||||   ::::::   :::::|||   :::|||   |||
      58 .....SerLysAsnSerAsnIleGluLeuTyrThrTyrSerProIleSer 72

     220 GGCAAAATCGCCGCTATTCACCGTGGC.........GAAAAGCGCGTACT 260
         |||   :::      |||:::  :::        :::   :::   ||
      73 GlyIleValGluLysIleTyrThrAlaAsnPheProAspGlyLysAsnLe 89

     261 TCAGTCAGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAAC 310
         |:::|||:::::|||      :::|||      :::|||:::   |||:
      89 uLysSerAlaLeuIleLysPheGlnGly......LysIleLysThrGluL 104

     311 GCTACGTACCTGAAGCGCTGGCAAAATTGAGCAGCGAAAAAGTGCGCCGC 360
         ::        |||      |||      |||   ||||||:::   :::
     104 ysThrProIleGluGluGlyAla......SerArgGluLysThrLeuGlu 118

     361 AACCTGATTCAATCAGGCTTA...TGGACTGCGCTTCGCACC......CG 401
         :::|||||||||   |||:::   |||          ::::::      ::
     119 LysLeuIleGlnLeuGlyIleProTrpPheAsnGluAsnSerLeuPheGl 135

     402 TCCGTTCAGCAAAATCCCTGCCGTAGATGCCGAGCCGTTCGCCATCTTCG 451
         :   ::::::|||         :::|||          :::   :::::::
     135 nPheIleAsnLysCysLysLysIleAsp......LysMetIleLeuLeuT 150

     452 TCAATGCGATGGAC......ACCAATCCGCTGGCTGCCGACCCTACGGTC 495
         ::|||:::   |||      ||||||          |||        ,
     150 hrAsnGlyLysAspAlaPheThrAsnIleSerGluAla..... 162

     496 ATCATCAAAGAAGCCGCCGAAGAC..... 519
         ::::::   |||      |||:::
     163 LeuMetAlaGluLysLeuGluGluIlePheSerGlyPheGlnIleIleAs 179

     520 .....TTCAAACGCGGCCTGTTGGTATTGAGCCGCCTGACCG 556
                 ||||||:::   :::::::::::|||:::      :
     179 pLysIlePheLysPheLysGluIleIleIleIleSerAsnLysAspL 196

     557 AA.............CGTAAAATCCATGTGTGTAAAGCAGCAGGCGCA 591
         ::              :::|||:::   :::   |||
     196 ysLeuLysLysGluPheGluLysLeuSerIlePheLysAsnArgLysIle 212

     592 GACGTGCCGTCT...GAAAATGCTGCCAATATCGAAACACATGAATTTGG 638
         :::   |||   |||||||||           :::|||||:::
     213 LysIleLysSerLeuGluAsnAlaTyrProTyrThrAsnHisGluMet.. 228

     639 CGGCCCGCATCCTGCCGGCTTGAGTGGCACGCACATTCATTTCATC.... 684
                                          :::|||||:::
     229 ..........................IleMetHisPheLeuTyrA 235

     685 .............GAGCCAGTCGGCGCGAATAAAACCGTGTGGACC 717
                     :::   ::::::   ::::::::::::
     235 snAsnLysAsnThrLysAspAspIleAsnProHisAsnAsnIleLeuLeu 251

     718 ATCAATTATCAAGACGTGATTGCTATCGGACGTTTGTTCGTAACAGGCCG 767
         |||   :::|||:::          :::   ::::::   ::::::::
     252 AlaAsnIleGluAspLeuTyrAsnAlaAsnLeuValIleLysAsnAsnAs 268

     768 TCTGAATACCGAGCGCGTGGTTGCCTTGGGCGGCCTGCAAGTCAACAAAC 817
         :         |||:::   ||||||:::::|||   :::      |||
     268 nProTyrLysGluLysPheValAlaIleAsnGlyAsnLysLysIleLysS 285

     818 CGCGCCTCTTGCGTACCGTTTTGGGTGCGAAGGTGTCTCAACTTACCGCC 867
           |||:::|||::::::   :::|||::::::   |||||||||
     285 erArgIleLeuLysValLysIleGlyThrSerPheSerGlnLeu...Ile 300

     868 GGCGAATTGGTTGAC 882
         :::|||   :::|||
     301 AsnGluLysIleAsp 305

seq_name: sp_invertebrate:O61228

seq_documentation_block:
ID   O61228      PRELIMINARY;      PRT;   810 AA.
AC   O61228;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   THR4 PROTEIN.
GN   THR4.
OS   Tenebrio molitor (Yellow mealworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC   Cucujiformia; Tenebrionidae; Tenebrio.
OX   NCBI_TaxID=7067;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99449601; PubMed=10518792;
RA   Mouillet J.F., Bousquet F., Sedano N., Alabouvette J., Nicolai M.,
RA   Zelus D., Laudet V., Delachambre J.;
RT   "Cloning and characterization of new orphan nuclear receptors and
RT   their developmental profiles during Tenebrio metamorphosis.";
RL   Eur. J. Biochem. 265:972-981(1999).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AJ005685; CAA06670.1; -.
DR   HSSP; P19793; 2NLL.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
```

```
KW   Zinc-finger.
SQ   SEQUENCE   810 AA;  89535 MW;  A55D987DCFC7F489 CRC64;


alignment_scores:
            Quality:  106.50             Length:    328
              Ratio:  0.696                Gaps:     13
 Percent Similarity:  46.646   Percent Identity: 22.561

alignment_block:
 US-09-303-518D-131/rev x O61228   ..

 Align seg 1/1  to: O61228  from: 1  to: 810

    1304 AGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAG 1255
         |||:::|||           |||:::::   |||:::   :::|||
      43 SerSerProAspLeuValSerProSerSerProLysMetSerGluAla.. 58

    1254 GTCTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTAT 1205
                  |||      :::|||   :::      :::::::::
      59 .......ValLeuAsnProProSerProAspSerThrProIleGlnIleH 73

    1204 CGCCG.................ACGATTAAATCGCGCAAAAGCAAGGTA 1173
           |||                      |||:::|||||   :::
      73 isProProGluProThrLeuAspLysIleSerSerArgLeuGluAspSer 89

    1172 GGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGATC......GG 1129
         |||   ::::::::::|||   |||   |||   :::            ||
      90 GlyValPheAspGlyGlyGlyThrAlaSerValIleArgSerLeuProGl 106

    1128 TACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCGTGAACTTGAAGAGTT 1079
         |:::   ::::::||||||||||||
     106 yValArgSerGlnSerProProLeu..................... 114

    1078 TGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTGTCC 1029
                                     |||                 |||
     115 ............................ArgProAlaProThrSerSer 121

    1028 GGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGAT 979
           |||         :::|||:::||||||      ||||||:::
     122 ProCysGlnProProArgProHisSerSerProGlyArgProAla..... 136

     978 AACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCTTGTG 929
                                                :::   :::|||
     137 ..............................AsnThrSerProValI 142

     928 CAATC...GCACCGTTCAATACCGAACCGGAAATCACGCGGTTGTCCGCG 882
           |||      |||      ::::::|||      :::|||   :::
     142 leIleHisAsnProThrIleAlaGlnProThrSerSerArgAlaLysLeu 158

     881 TCAACCAATTCGCCGGCGGTAAGTTGAGACACCTTCGCACCCAAAACGGT 832
         ||||||         :::   :::      :::   :::|||::::::||
     159 SerThrMetValThrCysGlnGluProThrSerSerSerProSerAlaVa 175

     831 ACGCAAGAGGCGCGGTTTGTTGACTTGCAGGCCGCCCAAGGCAACCACGC 782
         |   |||   ||||||:::::      :::|||   :::   |||
     175 lSerLysSerArgGlyValIleIleSerHisProThrSerLeuThrGlnS 192

     781 GCTCGGTATTCAGACGG......CCTGTTACGAACAAACGTCCGATAGCA 738
           :::::::::   :::         :::::::::   :::|||
     192 erGlnLeuTrpMetLysHisSerArgIleAsnGlyValLysProGluLeu 208

     737 ATC.........ACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGC 697
         |||          |||      |||      ::::::   |||   ::
     209 IleGlyGlyAsnPheSerGlyAlaLeuGlyHisTyrSerGluLeuLysSe 225

     696 GCCGACT......GGCTCGATGAAATGAATGTGCGTGCCACTCAAGCCGG 653
         :||||||      |||:::|||:::                 :::||||:
     225 rProThrProAlaGlyAlaMetGln.................ArgProS 236

     652 CAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATTGGCAGCATTTTCA 603
         :::::      |||   ::::::      :::::::::   :::   :::
     236 erSerAsnProProValArgGlnThrProThrValIleMetGlyGluAla 252

     602 GACGGCACGTCTGCGCCTGCTGCTTTACACACATGGATTTTACGTTCGGT 553
           |||                                           ||
     253 GlyGly.....................................Va 255

     552 CAGGCGGCTCAATACCAACAGGCCGCGTTTGAAGTCTTCGGCGGCTTCTT 503
         ||||   :::      :::::::|||   |||   :::|||   :::|||
     255 lArgThrMetIleTrpSerGlnProThrLeuGlyAlaSerProThrSerP 272

     502 TGATGATGACCGTAGGGTCGGCAGCCAGCGGATTGGTGTCCATCGCATTG 453
           :::   :::         |||:::::::::   :::|||   |||
     272 roValGluSerProHisHisAlaThrThrSerTrpAlaSerGlyAla... 287

     452 ACGAAGATGGCGAACGGCTCGGCATCTACGGCAGGGATTTTGCTGAAC.. 405
           :::|||:::|||   :::   |||:::|||   :::||||||||||
     288 ...AsnMetSerAsnThrGluGluSerAlaAlaGlnMetLeuLeuAsnLe 303

     404 .......GGACGGGTGCGAAGCGCAGTCCATAAG 378
                 |||:::||||||   |||   :::
     303 uGlyGlnAspArgLeuArgSerProValSerArg 314

seq_name: sp_human:Q96HA1

seq_documentation_block:
ID   Q96HA1      PRELIMINARY;      PRT;   984 AA.
AC   Q96HA1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:3792).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LYMPH, AND LYMPHOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC008794; AAH08794.1; -.
SQ   SEQUENCE   984 AA;  99036 MW;  36DEDB4419EEA8E6 CRC64;


alignment_scores:
            Quality:  106.50             Length:    520
              Ratio:  0.465                Gaps:     22
 Percent Similarity:  44.038   Percent Identity: 21.346

alignment_block:
 US-09-303-518D-131/rev x Q96HA1   ..

 Align seg 1/1  to: Q96HA1  from: 1  to: 984

    1310 CGCAACAGCGGGCCGTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAA 1261
         :::|||||:::      |||      ||||||::::::      :::||
      225 GlnAsnSerAsnSerGlnSer...ThrProGlySerSerGlyGlnArgLy 240

    1260 AGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAAAGCCTGCGCGCTGT 1211
         |   :::   :::          |||::::::   :::      :::|||
      240 sArgLysValGlnLeuLeuProSerArgArgGlyGluGlnLeuThrLeuP 257

    1210 CGGTATCGCCGACGATT.............................. 1194
                |||   :::
      257 roProProProGlnLeuGlyTyrSerIleThrAlaGluAspLeuAspLeu 273

    1193 ............................AAATCGCGCAAAAGCAAGGT 1174
                                     ::::::   :::::::|||
      274 GluLysLysAlaSerLeuGlnTrpPheAsnGlnAlaLeuGluAspLysSe 290

    1173 AGGCAGGATGTCCAACGGCATTACGCGCTCATAAGTGCCGATC....... 1131
                |||||::::::|||::::::      ||||||
      290 rAspAlaAlaSerAsnSerValThrGluThr...ProProIleThrGlnP 306
```

```
1130 ......................GGTACCATGGCGCGGTCGCCGCCGTTG 1104
                           ::::::         |||||||||
 306 roSerPheThrPheThrLeuProAlaAlaAlaProAlaSerProProThr 322

1103 ACGGCTGTCGTGAAC..............TTGAAGAGTTTGTTTTTTAG 1069
     :::   ::::::                 |||:::||||||
 323 SerLeuLeuAlaProSerThrAsnProLeuLeuGluSerLeuLysLysMe 339

1068 GAAATGGCCGAGAGTGGTGCGCGTGATGGAGTATTTGTCCGGCTGCGGCG 1019
      :::   |||      :::                 :::|||:::::::
 339 tGlnThrProProSerLeuProProCysProGluSerAlaGlyAlaAlaT 356

1018 CAACCCAGCCGAACAGC..............TCTTTGCTGCGGCCTTCT 984
     ::|||:::      |||             |||||||||   |||
 356 hrThrGluAlaLeuSerProProLysThrProSerLeuLeuProProLeu 372

 983 TCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCC 934
     ::::::::::::                                :::||
 373 GlyLeuSerGln............................SerGlyPr 379

 933 TTGTGCAATCGCACCG.........TTCAATACCGAACCGGAAATCACGC 893
     |   ::::::   |||         |||::::::::::|||      |||
 379 oProGlyLeuLeuProSerProSerPheAspSerLysProProThrThrL 396

 892 GGTTGTCC.........GCGTCAACCAATTCGCCGGCGGTAAGTTGAGAC 852
       |||:::         |||   :::      |||||||:::      |||
 396 euLeuGlyLeuIleProAlaProSerMetValProAlaThr......Asp 410

 851 ACCTTCGCACCCAAAACGGTACGCAAGAGGCGCGGTTTGTTGACTTGCAG 802
     |||   ||||||   |||||||:   :::            |||:::
 411 ThrLysAlaProProThrLeuGlnAlaGlu............ThrAlaTh 423

 801 GCCGCCCAAGGCAACCACG.................CGCTCGGTATTCA 770
         |||:::||||||:::                 ::::::   |||
 423 rLysProGlnAlaThrSerAlaProSerProAlaProLysGlnSerPheL 440

 769 GACGGCCTGTTACGAACAAACGTCCGATA..............GCAATC 735
           :::   |||      |||                   |||
 440 euPheGlyThrGlnAsnThrSerProSerSerProAlaAlaProAlaAla 456

 734 ACGTCTTGATAATTGATGGTCCACACGGTTTTATTCGCGCCGACTGGCTC 685
     :::|||         |||          ::::::   ||||||      ||
 457 SerSerAlaProProMetPheLysProIlePheThrAlaProProLysSe 473

 684 GATGAAATGAATGTGCGTGCCACTCAAGCCG..................... 654
     |   |||          :::|||      |||
 473 rGluLysGluGlyProThrProProGlyProSerValThrAlaThrAlaP 490

 654 ..................................................... 654

 490 roSerSerSerSerLeuProThrThrThrSerThrThrAlaProThrPhe 506

 653 ............GCAGGATGCGGGCCGCCAAATTCATGTGTTTCGATATT 616
                 ::::::   |||||||||   |||   |||   :::
 507 GlnProValPheSerSerMetGlyProProAlaSer...ValProLeuPr 522

 615 GGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTACACACATGGA 566
      |||   |||          |||::::::||||||:::      |||
 522 oAlaProPhePheLysGlnThrThrThrProAlaThrAlaProThr.... 537

 565 TTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGTTTGAAG... 519
                            |||:::   |||   :::
 538 ........................ThrThrAlaProLeuPheThrGly 545

 518 ...TCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAGCGG 472
        :::|||||||:::|||   :::      :::   |||||||:::   ::
 546 LeuAlaSerAlaThrSerAlaValAlaProIleThrSerAlaSerProSe 562

 471 ATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGC......TCGGCAT 428
     :       |||   :::          ::::::   |||      |||:::|
 562 rThrAspSerAlaSerLysProAlaPheGlyPheGlyIleAsnSerValS 579

 427 CTACGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAG 378

     ||:::::::::::::      :::      :::   :::|||:::   :::
 579 erSerSerSerValSerThrThrThrSerThrAlaThrAlaAlaSerGln 595

 377 CCTGATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAG 328
     |||                        |||:::      :::   :::::
 596 Pro..................PheLeuPheGlyAlaProGlnAlaSerAl 606

 327 CGCTTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTG......CCTT 284
     :||||||   |||          :::||||||   ::::::      |||
 606 aAlaSerPheThrProAlaMetGlySerIlePheGlnPheGlyLysProP 623

 283 CAACGGCAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATA 234
       :::      ||||||          |||   ||||||          :::
 623 roAlaLeuProThrThrThrThrValThrThrPheSerGlnSer...Leu 638

 233 GCGGCGATTTTGCCTGAAGCCGGCGCAGTAAATACTACG......CCCGG 190
       :::   :::|||   |||   :::   :::::::::         ||
 639 HisThrAlaValProThrAlaThrSerSerSerAlaAlaAspPheSerGl 655

 189 ATTCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCTTCACCTT 140
     ||||      :::      ::::::   |||      ::::::::::|||:
 655 yPheGlySerThrLeuAlaThrSerAlaProAlaThrSerSerGlnProT 672

 139 CCTTGATTTTCATCGAGGGGCGCATGCCGACATATTCTTCGCCAAGCAAC 90
     ::|||   |||   :::         ||||||::::::   |||   :::
 672 hrLeuThrPheSerAsnThrSerThrProThrPheAsnIleProPheGly 688

  89 GCGACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCC 40
     :::::::::            ||||||          ::::::      ||
 689 SerSerAlaLysSerProLeuProSerTyrProGlyAlaAsnProGlnPr 705

  39 CGCGATGGGC 30
     ||||:::|||
 705 oAlaPheGly 708
```

seq_name: sp_invertebrate:Q95SX4

seq_documentation_block:

```
ID   Q95SX4      PRELIMINARY;      PRT;   632 AA.
AC   Q95SX4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LD47460P.
GN   CG4453.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, CN BW SP;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY060443; AAL25482.1; -.
SQ   SEQUENCE   632 AA;  61593 MW;  781D83A42C6557E3 CRC64;
```

alignment_scores:

```
            Quality:  106.00              Length:    538
              Ratio:  0.453                 Gaps:     19
Percent Similarity:  43.494   Percent Identity: 20.260
```

alignment_block:

```
US-09-303-518D-131/rev x Q95SX4   ..

Align seg 1/1  to: Q95SX4  from: 1  to: 632
```

```
1343 CAGCCTTCCTTCTCAATGGTTTCCAGCACTTTGCGCAACAGCGGGCCGTA 1294
  46 LysProSerValGluThrThrThrThrProAlaProThrSerAlaGlnPh 62

1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
  62 eSerPheGlyPheGlyGlnSerAsnGlnGlyLysAspValAlaAspSerL 79

1243 CCAATTCCAAGCAACCCAAAGCCTGCGCGCTGTCGGTATCGCCGACGATT 1194
  79 ysLysThrGluAlaProLysThrPheMetPheGlyValSer..... 92

1193 AAATCGCGCAAAAGCAAGGTAGGCAGGATGTCCAACGGCATTACGCGCTC 1144
  93 LysValGluGluProLysThrValSerPheGlyThrGlyIleLysGluTh 109

1143 ATAAGTGCCGATCGGTACCATGGCGCGGTCGCCGCCGTTGACGGCTGTCG 1094
 109 rThrAlaThrSerSerThrGluAlaThrAlaProThrProAlaAlaAlaA 126

1093 TGAACTTGAAGAGTTTGTTTTTTAGGAAATGGCCGAGAGTGGTGCGCGTG 1044
 126 laProAlaProValGlnPheValPheLysAlaProThrThrAlaThrThr 142

1043 ATGGAGTATTTGTCC..... 1029
 143 AlaSerSerLeuThrThrThrIleSerThrThrSerAsnAlaProAlaLe 159

1028 ....GGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTT 983
 159 uGlyGlyPheSerPheGlyAlaProSerSerSerSerThrValSerSerS 176

 982 CGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCATGCGCGCCT 933
 176 erThrThrSerThr.....SerAlaAsnPro 184

 932 TGTGCAATCGCACCGTTC..... 915
 185 AlaAlaValLysProMetPheSerTrpSerGlyAlaGlySerAlaValSe 201

 914 .....AATACCGAACCG..... 903
 201 rSerThrSerSerSerGlnGlnProValAlaLysAlaProThrLeuGlyP 218

 902 .....GAAATCACGCGGTTGTCCGCGTCAACCAAT... 873
 218 heGlyValSerSerSerThrValThrThrThrThrThrSerThrLysVal 234

 872 .....TCGCCGGCGGTAAGTTGAGACACCTTCGCA..... 843
 235 PheAlaPheThrProAlaSerGlyLeuAspProAlaAlaAlaThrSerAl 251

 842 .....CCCAAAA 836
 251 aProAlaAlaGlyAlaGlyPheSerPheGlySerGlnSerLysProAlaT 268

 835 CGGTACGC.....AAGAGGCGCGGTTTGTTG 810
 268 hrThrGlnAsnThrGlyThrPhePhePheGlyGlnProThrAlaValAla 284

 809 ACTTGCAGGCCGCCCAAGGCAACCACGCGCTCGGTATTCAGACGGCCTGT 760
 285 ProAlaThrProThrAsnProSerValSerSerIlePheGlyAlaProAl 301

 759 TACGAACAAACGTCCGATAGCAATCACGTCTTGATAATTGATGGTCCACA 710
 301 aThrSerThrThrAlaSerThrSerValSerAlaThrThrSerThrSerT 318

 709 CGGTT.....TTATTCGCGCCGACTGGCTCGATGAAATGA 675
 318 hrAlaAsnAlaIleAlaSerSerPheAlaProThrSerThr..... 331

 674 ATGTGCGTGCCACTCAAGCCGGCAGGATGCGGGCCGCCAAATTCATGTGT 625
```

```
 332 .....ProGlnLeuPheGlyAsnTrpGlyGluLysLysThrAspLe 345

 624 TTCGATATTGGCAGCATTTTCAGACGGCACGTCTGCGCCTGCTGCTTTAC 575
 345 uThrThrPheGlyAlaSerSerGlySerGlyThrThrThrThrProSerP 362

 574 ACACATGGATTTTACGTTCGGTCAGGCGGCTCAATACCAACAGGCCGCGT 525
 362 heGlyTrp.....SerSerAsnGlyAspAla 370

 524 TTGAAGTCTTCGGCGGCTTCTTTGATGATGACCGTAGGGTCGGCAGCCAG 475
 371 AlaLysSerAsnSerAlaAla.....ValGlySerAlaAla.. 382

 474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCTA 425
 383 .....ValProSerSerSerAlaSerT 390

 424 CGGCAGGGATTTTGCTGAACGGACGGGTGCGAAGCGCAGTCCATAAGCCT 375
 390 hrMetAlaThrProIlePheGly..... 397

 374 GATTGAATCAGGTTGCGGCGCACTTTTTCGCTGCTCAATTTTGCCAGCGC 325
 398 .....SerSerSerMetPheGlyProSerSerSe 407

 324 TTCAGGTACGTAGCGTTCGAACTCGATTTCGTCGTTGCCT...TCAACGG 278
 407 rSerAsnAsnThrThrSerThrSerThrThrSerLeuProPheGlySerA 424

 277 CAATCACGACTGACTGAAGTACGCGCTTTTCGCCACGGTGAATAGCGGCG 228
 424 laAlaThrThr.....AlaAla 429

 227 ATTTTGCCTGAAGCCGGCGCAGTAAATACTACGCCCGGATTCTTTTTGTC 178
 430 ThrThrProAlaGlyGlyAsnAlaAlaLeuThrGlyLeuPheGlyAsnVa 446

 177 TTCAAACAGC.....ACTTGGCCTTTTT 155
 446 lGlyAsnSerLeuAlaGlyValGlyAlaProValAlaThrThrProAlaA 463

 154 TGACGGCTTCACCTTCCTTGATTTTCATCGAGGGGCGC..... 117
 463 laThrAlaAlaAlaProLeuThrAsnIlePheGlyAsnProThrProVal 479

 116 .....ATGCCGACATATTCTTCG.....CCAAGCAACGC 88
 480 AlaAlaAlaAlaProValPheGlySerGlySerThrIleProSerAlaGl 496

  87 GACTTCGGTAATGGCCGGGCCGTCATAAATGACTTGCTCCGGTCTGCCCG 38
 496 yPheGlyAlaProAlaAlaAlaAlaProLeuAlaAlaProAlaLeuProG 513

  37 CGATGGGCAGATTT 24
 513 lyAlaPheAsnPhe 517
```